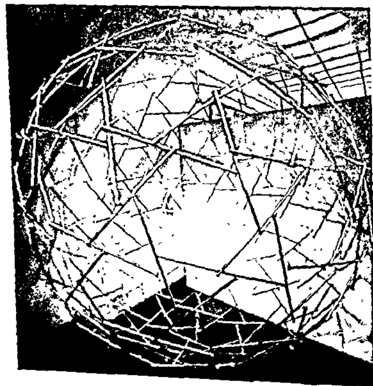
McGraw-Hill Encyclopedia

McGRAW-HILL BOOK COMPANY INC.

NEW YORK CHICAGO SAN FRANCISCO DALLAS TORONTO LONDON

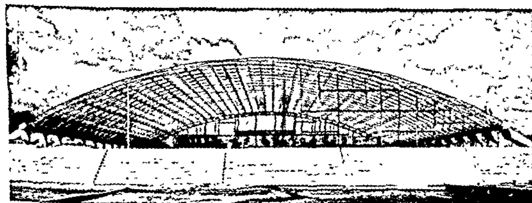


of Science and Technology

AN INTERNATIONAL REFERENCE WORK

IN FIFTEEN VOLUMES INCLUDING AN INDEX

VOLUME 5 ENT-FUS



(LEFT) A tensegrity sphere built according to the principles of tension integrity developed by R. Buckminster Fuller Tension forces in the structure are handled by thin wires. The continuous pull of the wires is resisted by the isolated discontinuous tubes which are in compression *(photograph by Hans Namuth)* (RIGHT) Lamella steel arches for the civic auditorium in Corpus Christi, Texas *(photograph by Sammy Gold)*.

Library of Congress Catalog Card Number: 60-11000

Suggestions to the Readers

The basic plan of the Encyclopedia is explained here in order to facilitate its use.

The subject matter of the various disciplines or branches of science and technology is organized systematically: a general article provides a broad survey of the field, and a number of separate articles, alphabetically arranged, cover its main subdivisions and more specific aspects.

Cross references guide the reader from the general articles to the other articles into which the subject is subdivided, and from these to articles on more highly specialized phases of the subject. The cross references—there are about 50,000 of them—are *printed in small capital letters so that they can be* easily recognized. By means of the cross references a reader may find his way from ELECTRICAL ENGINEERING, through ELECTRONICS and VACUUM TUBE, to ELECTRON MOTION IN VACUUM or ELECTRON EMISSION. Or, following another line of cross references, the reader would be led to ELECTRIC POWER SYSTEMS, TRANSMISSION LINES, ELECTROMAGNETIC WAVE, and so on.

increasingly complex and detailed considerations. A reader thus needs to proceed only as far as his inclinations and requirements dictate.

The Index, Volume 15, should be consulted to locate the discussion of topics covered in the Encyclopedia but not given in separate entries.

Every phylum, class, and order in the plant and animal kingdoms is allotted a separate article. Many of the more common families, genera, and species are covered either in one of the order articles or in a separate article under its own scientific or common name.

The adjectives electric and electrical are used in the following senses. Electric—containing, producing, arising from, actuated by, or carrying electricity, or capable of doing so; as, for instance, electric generator, electric motor, electric wiring. Electrical—related to, pertaining to, or associated with electricity, but not having its properties or characteristics; as, for example, electrical code, electrical engineering.

Words used as titles are, wherever possible, given in the singular to permit a consistent alphabetic arrangement. Titles are alphabetized by word and not by letter; for example,

Earth sciences
Earth tides
Earthmover
Earthquake

A word used as a noun precedes the same word used adjectivally; thus:

Mercury (element)
Mercury (planet)
Mercury battery

or

Circuit, electronic
Circuit breaker

Hyphened terms are alphabetized as single words; for example,

Animal virus
Animal-feed composition

Most of the longer articles contain bibliographies citing useful sources of further information. For additional bibliographical citations, the reader should refer to related articles (as indicated by the cross references in the article) Bibliographies are placed at the ends of articles or sometimes at the ends of major sections in long articles.

A list of initials and names of the contributors to the Encyclopedia is to be found in Volume 15. This list will permit quick identification of a contributor's initials after an article. Immediately following this list is a second list of encyclopedia contributors with their affiliations and the titles of articles each has written for the Encyclopedia.

McGraw-Hill Encyclopedia of Science and Technology

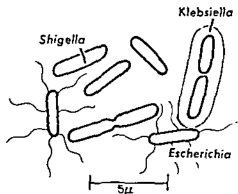
Enteric bacilli

Gram-negative, rod-shaped bacteria in the intestinal tract of man and animals. The normal flora are species in the following genera: *Escherichia*, *Aerobacter*, *Proteus*, *Paracolobacterium*, and *Pseudomonas*. The intestinal pathogens are species in the *Salmonella* and *Shigella*. Some of the normal flora like *Escherichia coli* and *Pseudomonas aeruginosa* are under certain conditions pathogenic, causing urinary tract and wound infections. Intestinal pathogens cause diseases like typhoid fever, paratyphoid fever, and bacillary dysentery. For taxonomy *see* AEROBACTER; ESCHERICHIA; KLEBSIELLA PNEUMONIAE; PARACOLON BACILLI; PROTEUS; PSEUDOMONAS AERUGINOSA; SALMONELLA; SHIGELLA; *see also* BACILLARY DYSENTERY; BACTERIOLOGY, MEDICAL; PARATYPHOID FEVER; PARATYPHOID GASTROENTERITIS; TYPHOID FEVER. [A.J.W.]

Enterobacteriaceae

A family of bacteria of the order Eubacteriales, consisting of straight, rod-shaped, gram-negative, nonsporeforming cells, motile with peritrichous flagella or nonmotile (see illustration). All members are facultative chemosynthetic heterotrophs obtaining energy by the oxidation of a variety of simple organic compounds or by the fermentation of sugars, and in some cases organic acids and polyalcohols. Most will grow in synthetic media with a single carbon source and an inorganic nitrogen source, although some species have requirements for specific amino acids. *See* BACTERIAL METABOLISM; BACTERIAL MOTILITY.

The family contains many species of direct significance in human affairs including important human and plant pathogens. This, combined with their widespread occurrence and ease of laboratory culture, has led to their extensive use as experimental organisms, and more is probably known about some of the enterobacteria, in particular *Escherichia coli*, than about any other bacteria. Increasing knowledge about the group has magnified rather than lessened the difficulties of classification, and nowhere do the unresolved basic problems of bacterial taxonomy stand out more clearly than



Some genera of the Enterobacteriaceae. (V. B. D. Skerman)

Key properties of the Enterobacteriaceae

Family and genus	Motility	Fermentation of lactose	Gas from sugars	Fermentation type	Other
Escherichieae					
Escherichia	+	+	+	A	
Aerobacter	±	+	+	S	
Klebsiella	−	+	+	A, S	Respiratory pathogens
Paracolobactrum	±	Delayed	+	A, S	
Alginobacter	+	Slow	+	A(?)	Alginic acid fermented
Erwinieae					
Erwinia	+	±	±	A, S(?)	Plant pathogens
Serratieae					
Serratia	+	+(?)	±	S, A(?)	Red pigment
Proteeae					
Proteus	+	−	+	A	Urease+
Salmonelleae					
Salmonella	+	−	+	A	Urease−
Shigella	−	−(or slow)	−	A	

+, Property present in most species; ±, property variable between species of the genus; −, property absent in most species; ?, unknown or uncertain; A, mixed acid; S, neutral solvent.

here. As a consequence, frequent revisions of the subdivisions of the family have occurred, and a stable classification is still to be attained. The existence of sexual reproductive mechanisms in bacteria was first shown for *E. coli* and very recently hybridization between *E. coli* and *Shigella* species has been demonstrated. Extension of such studies may ultimately lead to a sound taxonomy of the group. At present, the family is subdivided into 5 tribes and 10 genera on the basis of sugar fermentations and other metabolic differences, pigmentation, and pathogenicity. Morphology is not employed since all species are essentially identical in this respect. The family is divided into the tribes Escherichieae, Erwinieae, Serratieae, Proteeae, and Salmonelleae.

Fermentation reactions. The fermentation of sugars proceeds by two main pathways, resulting in one case in a mixture of organic acids, succinic, lactic, acetic, and formic, and in the other in neutral solvents, ethyl alcohol and butylene glycol, as the major end products. Appreciable gas evolution occurs in the former case, if the organism has the enzyme formic hydrogenlyase which catalyzes the cleavage of formic acid, yielding carbon dioxide and hydrogen in a 1:1 ratio. The latter pathway is characterized by a carbon dioxide–hydrogen ratio of much greater than 1, since carbon dioxide arises in the reactions leading to butylene glycol. The type of fermentation mechanism, as well as the particular sugars used, has been given paramount status in the subdivision of the family into smaller taxonomic units, and a variety of qualitative tests has been developed for the rapid characterization of the fermentation patterns.

Antigenic structure. The antigenic structure of certain groups of enterobacteria has been intensively studied. Antigenic analysis, initially undertaken for epidemiological reasons, has shown the existence of some 500 different *Salmonella* with respect to the combinations of somatic and flagellar antigens present, and these differences have served as the basis of various schemes of classification (*see* SALMONELLA). Although not as extensively investigated, it appears that the *Escherichia* are equally diverse antigenically and the *Shigella* somewhat less so. The presence of common antigens among species of different genera of the family is common and has been interpreted as showing a common phylogeny for the group.

Escherichieae. The tribe Escherichieae is defined by the ability to ferment lactose with the rapid production of acid and visible gas. The tribe is divided into five genera including the *Escherichia* that have the typical mixed acid fermentation and the *Aerobacter*, having the neutral solvent pathway. The former are a characteristic part of the intestinal flora of humans and animals, and may under certain circumstances be pathogenic to man. The latter, although frequently found in feces, are more typically associated with plant materials. The two genera comprise the bulk of the coliform group which serves as a pollution indicator for water and foods (*see* WATER ANALYSIS). The *Klebsiella* have been isolated mainly from respiratory infections in man, and apart from their pathogenic potentials cannot be distinguished from other Escherichieae. The *Paracolobactrum* differ from typical *Escherichia* or *Aerobacter* only in causing delayed or slow fermentation of lactose; and the *Alginobacter* have the single distinguishing feature of fermenting alginic acid. *See* AEROBACTER; ESCHERICHIA.

Erwinieae. The tribe Erwinieae, with the single genus *Erwinia*, comprises a group of plant pathogens that cause a variety of galls, wilts, and soft rots, and includes the causative agent of fire blight of pear, the first bacterial plant pathogen to be described. The production of pectinases is characteristic of the soft rot group. The tribe is not homogeneous with respect to fermentative pathways, both major mechanisms being present in the group, and gas may or may not be produced.

Serratieae. The tribe Serratieae, genus *Serratia*, is composed of widely distributed water and soil forms, most carrying out a neutral solvent type fermentation. They are characterized by the production of a bright orange to deep red pigment, prodigiosin. Bacteriological lore attributes the miracle of "bloody bread" to the *Serratia*. Until recently the pigment has been considered as a tripyrrole compound; at present there is some evidence in favor of a structure represented by a pyridine ring with two substituent pyrrole rings. Its function is unknown. It has antimicrobial properties and has been used experimentally in certain infections.

Proteeae. The tribe Proteeae, genus *Proteus*, includes forms common in feces and putrefying materials, which may be involved in intestinal and urinary tract infections in man. Characteristically, the representatives are motile, ferment dextrose but not lactose with gas production, and produce urease. Some strains are strongly agglutinated by the sera of patients having rickettsial infections of the typhus and tsutsugamushi fever groups, and are employed in laboratory diagnosis of these diseases (*see* RICKETTSIALES; RICKETTSIOSES). Although this cross-agglutination is due to the presence of common polysaccharide haptenes, little phylogenetic significance is attached to this phenomenon by most authorities. *See* PROTEUS.

Salmonelleae. The tribe Salmonelleae contains a variety of forms parasitic and pathogenic to man and animals, including the typhoid, paratyphoid, and bacillary dysentery bacteria (*see* BACILLARY DYSENTERY; PARATYPHOID FEVER; TYPHOID FEVER). The mixed acid fermentation is characteristic of the group, lactose is usually not fermented, and gas may or may not be produced. It is divided into two genera, the *Salmonella* and *Shigella*. The former are motile and produce gas from sugars (the typhoid organism excepted) while the latter are nonmotile and produce no gas. *See* SHIGELLA.

[S.C.R.]

Enterocoela

That section of the animal kingdom including the Echinodermata, Chaetognatha, Hemichordata, and Chordata. During embryonic development, the body

cavity or coelom is formed as pouches from the embryonic gut and not by a splitting of mesoderm as in the Schizocoela. This embryonic feature is one of several characteristics that lead some students to place the echinoderms and chaetognaths close to the chordates in animal classification. *See* SCHIZOCOELA; *see also* CHAETOGNATHA; CHORDATA; ECHINODERMATA; HEMICHORDATA. [T.I.S.]

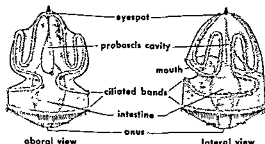
Enteropneusta

A class of the Hemichordata; known as the acorn worms or tongue worms. They are free-living solitary animals without exoskeleton. Numerous gill slits, a straight gut, and marked distinction between proboscis, collar, and trunk are characteristic of this class. Often, a free-swimming larva of the tornaria type occurs. Many of the dozen or so genera are widespread. Some species are locally abundant, but in general the group is uncommon. The distribution suggests that this is a group of great antiquity. Well-known genera include *Balanoglossus* and *Saccoglossus*.

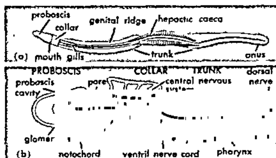
These excessively fragile, long, wormlike animals are usually burrowers in sand or mud. They feed upon the organic matter in the substratum which is passed through the gut. Some species throw up conspicuous castings. Many are described as smelling like iodoform; one is recorded as swimming; another is luminous.

The best-known group of the hemichordates, these animals are regarded as phylogenetically strategic. They are very simply organized and yet clearly related to and often treated as Chordata. A number of features suggest some affinity with the Echinodermata. The principal characters, in addition to the diagnostic ones above, are the following.

Two of the three coelomic segments open to the outside through special pores. The three segments can be homologized with the embryonic coelomic segments of echinoderms. Several organs are repeated but not in synchrony, and clear metamerism is absent. The gill slits are complex in structure, with a tongue bar and a skeletal rod of thickened basement membrane; synapticulae are often present. Frequently, these open to the outside via a pouch with a small external pore. A closed circulatory system is formed by splits in the basement membrane. There is a heart and pumping pericardium in the proboscis. A plexus of vessels in the



Tornaria larva. (*T. I. Storer and R. L. Usinger, General Zoology, 3d ed., McGraw-Hill, 1957*)



Saccoglossus, the tongue worm. (a) Dorsal view. (b) Median section of anterior portion, diagrammatic. (T. I. Storer and R. L. Usinger, General Zoology, 3d ed., McGraw-Hill, 1957)

posterior wall of the proboscis coelom, called the glomerulus, is supposed to be excretory. The stomocord is an anteriorly projecting diverticulum of the buccal cavity, hollow and with vacuolated cells. Between it and the external epithelium the basement membrane is greatly thickened into a nuchal skeleton. The very primitive nervous system consists of a nerve net generally distributed in the epithelia. It is thickened into cords in the middorsal and midventral lines in the trunk and is invaginated. In some genera, it is hollow in the collar region where no ventral but only the dorsal cord occurs. The dorsal cord is continuous with a plexus in the proboscis epithelium and a slight thickening thereof in the middorsal line. There are no organs of special sense, but some highly sensory epithelia are present on the proboscis. The gonads are simple sacs and numerous. They are arranged laterally in an irregular row in the trunk. In some genera they form dorsolateral folds, the genital alae or pleurae which end some distance behind the last gill slit. In this region, in some species, pouches of the gut bulge outward forming externally visible, so-called hepatic ceca. These occupy a short stretch, succeeded by a long unspecialized terminal trunk. The tornaria larva is large and marked by complex bands of cilia. It resembles externally and internally the auricularia larva of holothurians. There are biochemical similarities to both vertebrates and echinoderms. *See* CHORDATA; ECHINODERMATA; HEMICHORDATA; *see also* ENTEROCOELA. [T.H.B.]

Enterovirus

A family of animal viruses which include the 3 types of polio virus, the 25 types of Coxsackie virus, and 27 types of ECHO virus. They have been brought together as a family because they multiply chiefly in the alimentary tract, usually as a benign infection. When they invade other tissues, serious diseases may result, as when polio virus invades the spinal cord, or when a Coxsackie group B virus invades the heart muscle. The enteroviruses have a similar size, about 28 millimicrons in diameter. They are widespread in summer and fall in temperate climates, but are prevalent for longer periods

in tropical areas. *See* ANIMAL VIRUS; COXSACKIE VIRUS; ECHO VIRUS; POLIOMYELITIS. [J.L.M.]

Bibliography: E. Berger and J. L. Melnick (eds.), *Progress in Medical Virology*, vol. 1, 1958; *Nat. Foundation for Infantile Paralysis*, The enteroviruses, *Am. J. Public Health*, 47(12):1556–1566, 1957; T. M. Rivers and F. L. Horsfall, Jr. (eds.), *Viral and Rickettsial Infections of Man*, 3d ed., 1959.

Enterozoa

Animals with a digestive cavity or tract. The term excludes only the one-celled Protozoa, the peculiar Mesozoa, and the Parazoa or sponges that have many internal digestive cells but no cavity for that function. The form and extent of the digestive mechanism varies widely in other animal groups from a simple sac as in the *Coelenterata* to a diversified digestive tract in all higher groups. Tapeworms and a few other parasites lack a digestive tract. *See* MESOZOA; METAZOA; PARAZOA; PROTOZOA. [T.I.S.]

Enthalpy

For any system, that is, the part of space under discussion, enthalpy is the sum of the internal energy of the system plus the system's volume multiplied by the pressure exerted on the system by its surroundings. This may be expressed as $U + PV = H$. In this equation U is the system's internal energy, P the pressure of the surroundings on the system, V the system's volume, and H the enthalpy of the system. The sum of $U + PV$ is given the special symbol H primarily as a matter of convenience because this sum appears repeatedly in thermodynamic discussion. Consistent units must, of course, be used in expressing the terms in the above equation. Although each of the terms U and PV has the units of energy, enthalpy is not the energy of the system, but rather the sum of these two terms. Previously, enthalpy was referred to as total heat or heat content, but these terms are misleading and should be avoided. Enthalpy is, from the viewpoint of mathematics, a point function, as contrasted with heat and work, which are path functions. Point functions are usually much more readily handled in thermodynamic analysis than are path functions because the former depend only on the initial and final states of the system undergoing a change; they are independent of the paths or character of the change. Mathematically, the differential of a point function is a complete or perfect differential. (*See* CALCULUS, DIFFERENTIAL AND INTEGRAL; MATHEMATICS, MAXWELL'S EQUATIONS.) Because the absolute value of internal energy of even a simple system is usually unknown, recorded values of enthalpy are relative values measured above some convenient but arbitrarily chosen datum. Thus in the steam tables of Keenan and Keyes, the datum is liquid water at 32°F and under its own vapor pressure. At this state water is assumed to have an enthalpy equal to zero. Under this assumption the internal energy of water in this state is a negative quantity equal to PV. No complication is introduced by this fact, although visualization of negative energies of this kind may be disturbing to some. There is limited utility for absolute enthalpies because usually changes in enthalpy are of most interest. It is instructive to examine the utility of the enthalpy function in terms of some simple but important thermodynamic processes.

The first law of thermodynamics is merely a statement of the law of conservation of energy. The first law alone indicates that:

1. For a chemical reaction carried out at constant pressure and temperature and in the absence of changes in kinetic, potential, and similar energies, and with no work performed except that resulting from keeping the pressure constant as the volume changes, the change in enthalpy of the system (the material taking part in the chemical reaction) is numerically equal to the heat that must be transferred to maintain the above-mentioned conditions. This heat is often loosely referred to as the heat of reaction. More properly, it is the enthalpy change for the reaction.

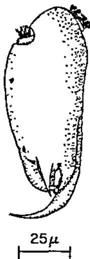
2. So-called heat balances on heat exchangers, furnaces, and similar industrial equipment that operate under steady flow conditions are really enthalpy balances.

3. The work developed in a steadily running adiabatic engine or turbine is equivalent to the enthalpy change of the steam or other fluid passing through the engine.

4. The adiabatic, irreversible, steady flow of a stream of materials through a porous plug or a partially opened valve under circumstances where difference in kinetic energy between entering and leaving streams is negligible (a Joule-Thomson process) results in no change in enthalpy of the flowing stream. Although no change in enthalpy results from this process, there is a loss in the energy available for doing work as a result of the Joule-Thomson flow. *See* ENTROPY; THERMODYNAMIC PROCESSES. For change in enthalpy with pressure or temperature, *see* THERMODYNAMIC PRINCIPLES. [H.C.W.]

Entodiniomorphida

An order of the Spirotricha. These are strikingly different-looking ciliates, covered with a smooth, firm pellicle. They are devoid of external ciliature except for the adoral zone of membranelles and, occasionally, one or two other tufts or zones of specialized cilia. Internal organization of the body is very specialized and complex. These organisms are considered to be highly evolved. Entodiniomorphids occur exclusively as endocommensals of herbivorous mammals, either in the rumen and reticulum of ruminants or in the colon of certain higher mammals. The ophryoscolecids, which comprise the majority of species in this order, are found in fantastic abundance in their hosts; for example, there have been estimations of as many as 10,000,000,000 per cow. Herbivores can survive free of their harmless protozoan guests, but the reverse



Ophryoscolex, an example of an entodiniomorphid.

is not true. *Epidinium*, *Ophryoscolex* (see illustration), and *Entodinium* occur in ruminants. *Troglodytella* is found in the colon of anthropoid apes. *See* SPIROTRICHA. [J.O.C.]

Entomology

The broad field of the biological sciences which treats all aspects of the study of insects. Insects such as *Drosophila* and *Habrobracon* have been used in genetic studies that have led to many significant discoveries. Insect physiology has provided many basic data in the fields of muscle physiology, endocrinology, and behavior.

Because of the large number of insect species and their economic significance, many specialized areas of entomology have developed. Numerous insects are beneficial to man. They are necessary for flower pollination upon which much of the fruit crop depends. Many important plants are attacked by fungal, viral, and bacterial diseases that are transmitted by insects. Insect parasites have been studied as a method of controlling destructive insects. The experimental animals have, naturally, been insects. Such studies form the basis of economic entomology.

The specialized field of medical entomology is concerned with insects that are vectors for numerous diseases and parasitic infestations of man and domestic animals. This is actually an aspect of economic entomology, and is of importance to public health officials studying epidemics in which insects are involved. *See* ENTOMOLOGY, ECONOMIC; INSECTA; INSECTICIDE. [C.B.C.]

Entomology, economic

Insects affect man in both destructive and helpful ways. They cause enormous economic losses each year. A major loss is the destruction or spoilage of food materials, occurring from the fields to the grocer's shelves. Another loss is the destruction of lumber, from the forest to the finished product, even after it is in use. A third loss is caused by the phytophagous species which feed upon vegetation valuable to man either as food or as raw material for manufacturing.

Insects are important in the transmission of disease, serving as vectors for organisms which cause many diseases of plants and animals, including man. Some of the most serious of these human diseases are typhus, bubonic plague, malaria, yellow fever, and African sleeping sickness. Transmission of many viral, bacterial, fungal, and protozoan diseases of plants can be attributed to insects.

On the other hand, insects contribute much to man's welfare. Various insect products, such as dyes, waxes, honey, and silk, although now superseded in large part by synthetics, are still of genuine value. Less tangible, but more important, is the role of insects in the pollenization of flowers, upon which fruit crops largely depend.

There is also the fact that the balance of nature is quite consistently maintained by the hosts of predaceous and parasitic insects which prey upon harmful species. Without these, human life could scarcely exist. In all, insects are at least as beneficial to man as they are harmful.

Lepidoptera. These are the butterflies and moths, which have a complete metamorphosis; the larvae are familiar as caterpillars. The pupae are frequently concealed within a cocoon spun of silk. The cocoons of certain species of moth provide the silk of commerce, the silkworm being the foremost of these species. The value of its products runs into many millions of dollars a year. On the other hand, the destructiveness of the larvae of many forms perhaps runs into even more millions. *See* LEPIDOPTERA.

Hymenoptera. This order includes the bees, wasps, ants, and many other forms without common names. A large group of insects are called bees, which are for the most part solitary forms, all intimately concerned with the pollenization of flowers. The most familiar species is the common honey bee, one of the relatively few insects directly beneficial to man by reason of its wax and honey. It is also notable because of the high order of social development. Besides these forms, the most important, and also physically the smallest, are species that in their larval stages are parasitic in the bodies of various larvae of other insects. These have, for the most part, no common names. They are indirectly beneficial to man because of the control exerted by their parasitism upon many harmful insects. In fact, it might be said that there is a hymenopteran parasitic species for every other species of insect. *See* HYMENOPTERA.

Coleoptera. These are the beetles, which have a tremendous variety of form and which feed upon almost every type of organic material. Among them, the weevils are probably the most notorious pests. The cotton boll weevil has cost the cotton growers of the United States millions of dollars. There is a host of other species important to man because of their destruction of cultivated crops, forest materials, and stored products. *See* COLEOPTERA.

Diptera. These true flies are important vectors of disease, so much so that a special field of study, medical entomology, has evolved. Although other orders of insects are included in medical entomology the predominant species are found in the order Diptera, the principal group involved in disease transmission. They present an enormous diversity of life habits. The mosquitoes, in their larval stages, live exclusively in free but quiet water. The adults, which sometimes occur in enormous swarms, have the mouthparts formed for piercing the skin of animals and thereby become at least a nuisance to man. However, their greatest importance comes from their association with disease-producing organisms present in the blood of animals. The mosquitoes pick up such organisms with their food and later transfer them to other animals. Thus certain species transmit malaria, yellow fever, and other diseases of both man and animals. This capacity to transmit disease occurs in numerous other forms as well.

Some flies live as external parasites in the hair of mammals or in the feathers of birds. They feed upon the blood of their hosts and, in birds, are known to transmit certain diseases. In this group, the larva is retained within the body of the female until it is ready to pupate; these forms are called the Diptera pupipara. In some species, like *Gasterophilus intestinalis*, the horse botfly, the larvae live in the alimentary canal of the horse. Others live within the flesh, sometimes forming tumors. One such species, *Wohlfahrtia vigil*, a fleshfly, occurs on man in the American tropics, causing a severe cutaneous myiasis of the face and neck in infants when it deposits its larvae on exposed skin. *See* MYIASIS.

Many of the species live upon the juices of plants. Some of these have great economic importance because of the damage done to crops, an example being the notorious Hessian fly of wheat. Some flies are regarded as beneficial to man because they either prey upon other insects or are parasitic within them, in their larval stages. *See* DIPTERA.

The fleas, or Siphonaptera, live as parasites in the hair of mammals and the feathers of birds. They transfer readily from one host to another and are famous for their leaping powers. The habits of the fleas create the capacity to transmit disease, and apart from being a nuisance, they are chiefly known as the principal carriers of bubonic plague, or "black death." Fleas may serve as the [illegible]
at[illegible]
Se[illegible]

[illegible] of the Orthoptera are the grasshoppers, known historically from the migratory forms, or locusts. When they are adult and have reached a *certain* abundance in numbers, they migrate in tremendous swarms. Such swarms have been recorded from the earliest times and still occur in some parts of the world, notably in North Africa and Argentina. In earlier times, such swarms occurred in the United States. Coming from their breeding grounds somewhere in the northwestern mountain states, they reached almost as far eastward as the Mississippi River and the borders of Oklahoma, causing tremendous losses to cultivated crops. More limited migrations from uncultivated lands to cultivated areas are still frequently reported in the United States. Other species are the mantids, which are the only strictly predaceous forms. They are beneficial to man, destroying other insects for food. *See* ORTHOPTERA.

Dermaptera. The Dermaptera or earwigs are similar in form to the Orthoptera. A few species are destructive to gardens, especially the European earwig, *Forficula auricularia*. *See* DERMAPTERA.

Hemiptera. This order comprises two groups, the Heteroptera and the Homoptera, now generally regarded as distinct orders. All the members of both groups are plant feeders except a comparatively small number which feed upon the blood of other insects and sometimes on that of birds and mammals.

Perhaps the most notorious of the Hemiptera is the chinch bug. This insect feeds on grasses, especially corn, and ranges from Central America to Canada. It has caused losses amounting to millions of dollars in the United States.

The harlequin cabbage bug feeds upon plants of the cabbage family and sometimes causes large losses. The squash bug is an important pest of garden plants, and many others are of significance to gardening and agriculture.

One family, the Cimicidae, or bedbugs, which is a group of less than 100 species, has long been known as an unwelcome companion of man. Only two species occur on man, the remainder on birds and bats. It infests the dwelling places of these animals, coming out to attack them; in all its stages, it lives upon their blood. It is not known to carry any disease of man, although it often is suspected of doing so.

Another family, the Reduviidae, known as the assassin or kissing bugs, feeds normally upon the body juices of other insects. A few species feed on mammalian blood. Certain of these forms transmit Chagas' disease (American trypanosomiasis) in South America. *See* HEMIPTERA.

Homoptera. The most familiar of the Homoptera are the aphids, which cause enormous losses to garden crops. Another group, the coccids or scale insects, contains enemies of fruit trees and is noteworthy because it includes a number of species that are or have been of high economic value to man. Certain dyes are obtainable from a number of coccids. Shellac, derived from the lac insect, is used for varnishes and similar products. *See* HOMOPTERA; SHELLAC.

Anoplura. These are the sucking lice that live solely in the hair of mammals and feed upon their blood. They have the mouthparts peculiarly developed for piercing the skin. The eggs are attached to the hairs and the nymphs feed in the same manner as the adults. Two species occur on man and at times both become nuisances because of their num-

bers. One species, the head and body louse, is responsible for the transmission of diseases, including typhus, as well as relapsing fever and trench fever. *See* ANOPLURA.

Thysanoptera. Some species of Thysanoptera, or thrips, are of considerable economic importance because of their damage to vegetation.

Isoptera. These are the termites, or white ants. Their food is dead wood and colonies are commonly formed within wooden material. This group damages or destroys houses and similar structures. In some parts of the United States the losses from termite damage reach a considerable figure. *See* ISOPTERA.

Psocoptera. The book, bark, and dust lice constitute a rather small group of insects which seem to live primarily upon the spores of fungi on the bark of trees. They are relatively small in size. A few species enter houses. They are important pests in herbaria, museums, libraries, and granaries. *See* PSOCOPTERA.

Thysanura. These forms, commonly called bristletails, include some which enter houses, as the silverfish. They live upon materials contained in paper or in the bindings of books and are at times destructive. *See* THYSANURA.

Control of insect pests is based on the use of various insecticides applied regularly during spraying schedules for fruit trees and other vegetation. A knowledge of insect physiology is an important control factor in the use and type of insecticide employed. *See* INSECT PHYSIOLOGY; INSECTA; INSECTICIDE. [C.F.F.]

Bibliography: W. B. Herms, *Medical Entomology*, 4th ed., 1950; J. G. Leach, *Insect Transmission of Plant Diseases*, 1940; C. L. Metcalf, W. P. Flint, and R. L. Metcalf, *Destruction and Useful Insects*, 3d ed., 1951.

Entoprocta

A group of Recent, aquatic, sessile, soft-bodied, tiny (less than 1 cm tall) invertebrate animals of controversial systematic position. Originally classed as zoophytes, whose individuals were called polyps, they are now considered by some authorities to be an independent phylum, by others a subphylum, class, or subclass along with or under the Bryozoa.

In 1869 H. Nitsche split the Bryozoa into two groups, Ectoprocta and Entoprocta, on the basis of the location of the anal opening In Ectoprocta, the anus opens outside the row of tentacles which surround the mouth. In Entoprocta, both mouth and anus open within the circlet of tentacles.

Entoprocta were renamed Calyssozoa by A. H. Clark in 1921 and called Kamptozoa by C. Cori in 1929, but these synonyms are not widely accepted.

Most Entoprocta conform to the basic structural pattern of a soft knoblike or bell-shaped tentacled head, or calyx, borne on the tip of a narrower stalk.

Taxonomy. Entoprocta comprise 3 families, the Loxosomatidae, Pedicellinidae, and Urnatellidae, 12 genera, and about 90 species. Families differ mainly as to mode of budding, growth habit, anat-

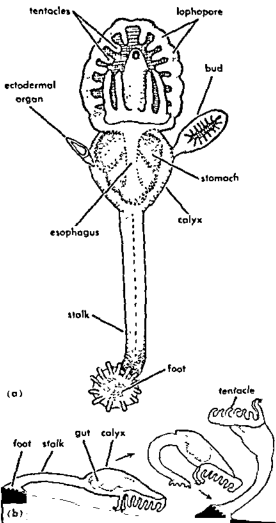
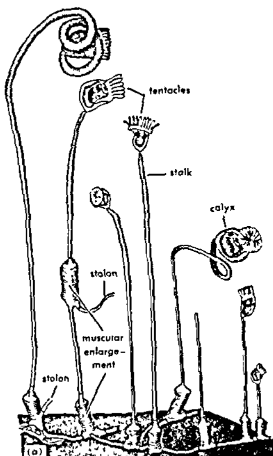


Fig 1. *Loxosoma saltans.* (a) Anterior aspect. (b) In "leap frog" locomotion. (*Modified after R Assheton, 1912*)

omy of the individual's stalk, and the individual's ability to shed or lose its calyx.

Loxosomatidae. This family contains about 58 solitary, marine species of the genera *Loxocalyx, Loxomespilon, Loxosoma,* and *Loxosomella* The calyx of each is attached to the substratum by a long or short, poorly defined, unpartitioned extension, the stalk, which in some species may have a pedal gland and attachment disk. No septum separates the calyx body cavity from that of the stalk. Calyces are not deciduous. Buds arise from the sides of the calyx.

Pedicellinidae. This entoproct family contains about 30 colonial, marine, or brackish-water species of the genera *Barentsia, Chitaspis, Coriella, Loxosomatoides, Myosoma, Pedicellina,* and *Pseudopedicellina* Colonies attach to the substratum by slender, partitioned, adherent stolons from which stalked individuals are budded at well-separated intervals. A septum separates the pseudocoel of the sharply defined deciduous calyx from the stalk



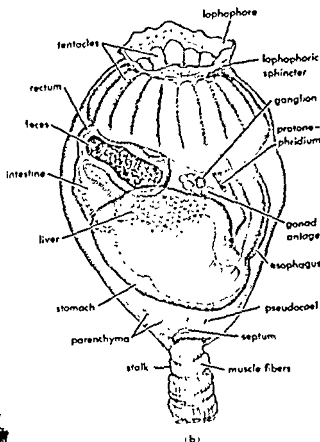


Fig. 2. *Barentsia laxa.* (a) Habit sketch. (b) Calyx.

Buds arise from the stolon but not from the calyx. Decapitated stalks can regenerate new calyces.

Urnatellidae. This family contains one genus, *Urnatella*, and two colonial, fresh-water species. Colonies have few individuals. One to six well-chitinized primary stalks arise from a flattened basal disk, which in the Indian species is segmented. Primary stalks have a variable number, 1–18, of segments between which are darker constricted nodes traversed by incomplete septa. Secondary stalks bud from primary stalk segments. A calyx develops on each new stalk tip. Stalks lose their calyces under adverse environmental conditions but can regenerate new ones.

Morphology. A lophophore, or fold of skin, bears on its inner surface a row of 8–36 partially ciliated tentacles which encircles the calyx or entoproct head. The tentacle number and degree of tentacular crown slant across the calyx differ in the various species. The entoproct crown can be rolled inward over the mouth and anus but is not retractile into the calyx interior, thus differing from the completely retractile ectoproct crown.

The transparency of the calyx and stalk varies with age and species. In *Barentsia*, the calyx is fairly transparent; in *Chitaspis*, which has a thickened cuticular shield, it is less so. The wall consists of a cuticle, an underlying epidermis, and some smooth muscle fibers.

The calyx contains the main ganglia and viscera which consist of a broad looped digestive tract, paired gonads and gonoducts, brood chamber (in some species) for developing larvae, and a pair of protonephridia. Protonephridia are excretory organs consisting of a flame cell bulb and several cells which presumably function as athrocytes. The gonoducts, nephridial tubes, mouth, and anus open to the outside within the circlet of tentacles.

The space between body wall and viscera is the pseudocoel. It contains a slight amount of very diffuse parenchyma consisting of mesenchymelike cells. *See* PSEUDOCOELOMATA.

The adult central nervous system consists of a large bilobed ganglion located in the calyx pseudocoel between the esophagus, stomach, and intestine. C. Cori, in 1929, showed five pairs of nerve trunks issuing from the *Pedicellina* subesophageal ganglion. Three pairs led to the tentacles, the fourth pair to the body wall, and the fifth pair to the stalk. S. Harmer, in 1885, showed tactile sense cells in the tentacles and a pair of lateral sense organs in the *Loxosoma* calyx. W. Hilton, in 1923,

found a peripheral nerve net in *Barentsia* stalk enlargements. However, the peripheral nervous system is still incompletely known.

The stalk supporting the calyx may have some of the following characteristics or appearances, depending upon the species: long to short; rigid to flexible; thick to slender; tapering, straight, moniliform, or with one or more muscular enlargements; undivided by septa, or segmented by one or more transverse septa; continuous with calyx pseudocoel, or cut off from calyx pseudocoel by septum; without a pedal gland or provided with a temporary or permanent pedal gland; stored food granules; special sensory, glandular, or amebocytic cells. Smooth muscle fibers occur in all entoproct stalks.

Adults of colonial species are sessile but can nod their heads, wave their stalks, flick or curl their tentacles, and tighten the lophophoric sphincter, because smooth muscle fibers are present in the stalk, calyx, lophophore and tentacles.

Adult solitary Loxosomatidae can change their location by creeping. Their detached buds may even swim a little, for short periods.

Reproduction. Entoprocts reproduce both sexually and by budding. Some species are hermaphroditic, some dioecious, and some protandric. Sometimes entire colonies are of one sex.

The gonads are a pair of testes and a pair of ovaries. Fertilization is internal. In *Pedicellina*, eosinophilic glands secrete a membrane about the egg for the egg's subsequent attachment to the em-

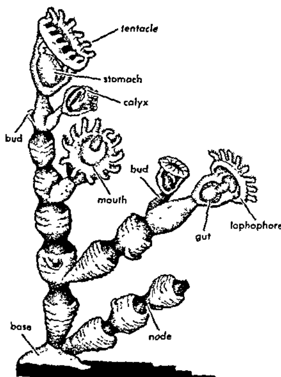


Fig. 3. *Urnatella gracilis*, habit sketch.

bryophore, a specialized vestibular epithelial thickening in the female calyx brood pouch. Cleavage is spiral. A coeloblastula appears by the 67-cell stage, and a gastrula by the 90-cell stage. This blastula has a slitlike, instead of rounded, blastocoel. Ciliated larvae are incubated in the brood chamber. Upon release, they creep or swim about for several hours before permanent attachment and metamorphosis.

Habitat. Entoprocts attach to rocks, algae, ascidians, sea anemones, corals, crinoids, crustacea, ectoprocta, hydroids, sponges, and worm tubes. Some Loxosomatidae also attach to gephyrean and polychete worms; however, the relationship is not obligatory, but merely accidental or commensal.

Entoprocts appear in all oceans, from polar to tropic regions, from shore level to a depth of 150 fathoms. See BRYOZOA. [M.D.RO.]

Bibliography: L. H. Hyman, *The Invertebrates*, vol. 3, 1951; M. Prenant and G. Bobin, *Faune de France: Bryozoaires*, vol. 60, 1956.

Entropy

A measure of the heat that flows during a thermodynamic process or of the information that is transmitted during a communication process. As a thermodynamic function, entropy provides criteria for spontaneous chemical charges and chemical equilibrium. In a heat engine, entropy is inversely proportional to the absolute temperature of a working substance during a particular process. An over-all increase in entropy measures a loss in ability to do work. See THERMODYNAMIC PROCESSES; THERMODYNAMICS (CHEMICAL).

Physical nature. Heat is a form of energy, but temperature alone is no more a measure of heat than is pressure a measure of work. Temperature is the intensity factor, it is perceptible to the senses. Entropy is the second factor, a distribution factor that bears a relation to heat and temperature analogous to that which volume bears to work and pressure. Just as volume is not sensible directly but only as a relation of three mutually perpendicular extensions, so entropy is not sensible directly but only by inference from several other observations.

The second law of thermodynamics requires the recognition of an extensive property of material systems to which the name entropy and symbol S are assigned. Entropy has a characteristic value for each state of the system. For a closed system (no transfer of matter across its boundaries) the change dS in entropy for any differential reversible process (in which the system is at all times only infinitesimally removed from a state of equilibrium with its surroundings) is equal to the heat δq absorbed from the surroundings divided by T, the absolute Kelvin or thermodynamic temperature at the system (see HEAT; TEMPERATURE). In contrast, if the process is irreversible (which it will be if it is of the natural, spontaneous type) change dS is greater than $\delta q/T$. Thus, although entropy is a state function and as such its change in any proc-

ess depends only on initial and final states of the system, the calculation of the entropy change must be made for a reversible path between the two states, because only then is $dS = \delta q/T$.

For a completely isolated system, entropy must either remain constant or increase, because δq is identically zero. Entropy always increases as a consequence of any irreversible process, however. Although the entropy of a particular system may actually decrease in an appropriate spontaneous process this system plus that portion of the surroundings which is in any way affected by the process together constitute a composite isolated system which experiences an entropy increase because of the irreversible process occurring within it. Entropy, unlike energy, thus is not conserved in nature, and the rate of production of entropy serves as the index of irreversibility.

Thermodynamic equilibrium. For a closed isolated system the increase of entropy accompanying spontaneous processes can continue only until the entropy has reached a maximum corresponding to attainment of a state of stable equilibrium. If the system does not interact with its surroundings, the internal energy and volume of the system must remain constant. If all the constituents are independent, mass M_i for each component i must individually stay constant. The basic thermodynamic criterion for stable equilibrium may then be written

$$dS = 0 \qquad E, V, \text{ all } M_i \text{ constant}$$

for a differential variation away from the equilibrium state, where S is entropy, E is internal energy, V is volume, and M_i is mass of component i. This relation is equivalent to the more convenient one

$$dE = 0 \qquad S, V, \text{ all } M_i \text{ constant}$$

The restriction to a system of independent components only may be eliminated by taking into account the mass balance constraints imposed by the stoichiometry of the equations of any reactions which can take place. Although stated for an isolated system, the above criteria are of general validity, because if a system is at equilibrium under any circumstances, the equilibrium will not be destroyed by placing a rigid adiabatic shield around the system.

Degradation of energy. The general tendency of entropy to increase inevitably leads to a conversion of energy into a form that is increasingly difficult to convert into work.

This increase in entropy for a system is associated in general with a shift to a state of lesser order at the molecular level. An example is the entropy increase accompanying the interdiffusion of two pure gases to form a solution. At the molecular level, the random motion of the molecules in the [illegible]

Statistical thermodynamics. Consider for example a gaseous system of fixed internal energy E and volume V, containing N identical molecules the interactions between which are negligible except as required to permit the assembly to reach equilibrium. The stationary energy states and energy levels for any molecule then correspond to those characteristic of the completely isolated molecule.

A quantum state of the whole system is defined by specifying the quantum state for each constituent molecule, with the reservation that states of the whole system which differ only by permutations of the indistinguishable subsystems among their individual states are not counted as different. A distribution of molecules over their possible energy states is made by assignment of n_i molecules to each permitted energy value ϵ_i, subject to the restrictions that total energy and total number of molecules are fixed:

$$\sum_i n_i = N \qquad \sum_i n_i\epsilon_i = E$$

If the degree of degeneracy is known for each energy level ϵ_i for a single molecule (that is, if the number of different quantum mechanical states consistent with energy ϵ_i are known for a molecule), the total number of quantum states for the total system consistent with the given distribution can be calculated. This total number will vary for different distributions, but will have a maximum value ω_0 for some particular distribution D_0. If now each of these quantum states is assigned an equal *a priori* probability of occurrence, distribution D_0 will become the most probable distribution and the corresponding state of the system the expected thermodynamic state.

There may then be defined a quantity S by the relation

$$S = k \ln \omega_0 \qquad k \text{ a constant}$$

It may be shown that S is an extensive property of the system, that it increases in any spontaneous process at constant E and V, and in general exhibits all the properties associated with thermodynamic entropy S. Finally, S becomes identical to S if constant k is set equal to the Boltzmann constant, which is the gas constant R divided by Avogadro's number. This relation is basic in statistical thermodynamics, which proceeds to compute the thermodynamic properties of the system as a whole from the properties of the individual units forming it.

For a perfect crystalline solid the value of ω_0 may often be considered to approach unity as the temperature approaches absolute zero. Entropy S then will approach zero at 0°K; this provides a statistical interpretation of the third law of thermodynamics. For ordinary chemical reactions, the contributions to the entropy from isotope mixing and nuclear spin degeneracy are ignored because they cancel out in any calculation of entropy changes. [P.B.]

Unavailable energy. In a system subjected to any reversible cyclical process, a summation around the cycle of all heat effects divided by the thermodynamic temperatures at which they took place is zero for the system and zero for its surroundings. Because the cyclical integral of $\int dQ_R/T$ is zero, the integral $\int dQ_R/T$ over any portion of the cycle must equal the negative of $\int dQ_R/T$ around the rest of the cycle. For any reversible process, the value of the integral for the system or for its surroundings is independent of the path; therefore the integral, which is the mathematical definition of the change in entropy, is a point function dependent only on initial and final states.

Symbol S designates total entropy of a substance, s designates specific entropy of a unit mass of a substance, and ΔS designates the change in entropy in going from state A to state B, so that $\Delta S = S_B - S_A$. The net entropy change for a system and its surroundings is always zero for reversible changes, and this sum for any irreversible change is always positive.

In the differential equation $T\,dS = dQ_R$, the left-hand member is the reversible heat effect as a result of a change. Thus, areas under curves on T-s diagrams are reversible heat effects, just as areas under curves on P-v diagrams are reversible mechanical work effects (see THERMODYNAMIC CYCLE). Net entropy increase (the sum of the entropy increase for the system plus that for the surroundings), when multiplied by the lowest temperature available for heat discard, gives the amount of energy that had previously been available for doing work before the change took place but now, because of the actual process followed, is irrecoverably lost as regards ability to do work in the particular process. In this respect, increases in net entropy are indirect measures of irreversibility or loss in ability to do work, unless a lower discard temperature can be found. All actual processes (some life processes perhaps excluded) are irreversible to a degree, and so the carrying out of all such actual processes results in an increase in the total entropy of the universe.

It would appear that, excluding some still unforeseeable reversing techniques or a source of energy entering our universe from beyond, the energy available for doing work must be continually decreasing; the universe must be running down. Conceivably, this trend could be reversed by life processes.

Relative entropies, values above the datum chosen, are reported in steam and similar tables of properties. At the datum condition, the substance for which the table is calculated is assumed to have zero entropy.

Absolute entropy. By measuring heat capacities down to as near absolute zero as is experimentally possible, taking proper cognizance of all entropy effects due to phase changes such as vaporization or crystallization, and taking the necessary precautions to exclude chemical reaction, absolute entropy can be determined for an element or a compound. The value of $\int dQ_R/T$ is then evaluated from the temperature at which the absolute entropy is desired down to the measurements made at the lowest experimentally obtainable temperature. From this temperature to absolute zero, usually a short range, various extrapolation methods are employed. Finally, at absolute zero, the entropy, if any, at that temperature is determined and added to the entropy value already found. The total is the absolute entropy. *See* THERMODYNAMICS (CHEMICAL).

A table of values of absolute entropies, together with a table of enthalpies (heats of formation), makes possible determination of equilibrium constants through the relation $-RT \ln K_a = \Delta H^0 - T\,\Delta S^0$, without recourse to experimentally determined equilibria.

Absolute entropies may be determined from spectroscopic data, although for complicated molecules the calculation is involved. A check of spectroscopically determined values against those determined by low-temperature calorimetric procedures is useful in detecting errors in either type of calculation.

There is a close parallel between entropy and probability, and, broadly speaking, entropy is a measure of the probability that a given system will exist in a certain state, the least-probable states represent those of lowest entropy. *See* BOLTZMANN STATISTICS; STATISTICAL MECHANICS. [H.C.W.]

Measure of information. This probability characteristic of entropy leads to its use in communication theory as a measure of information. The absence of information about a situation is equivalent to an uncertainty associated with the nature of the situation. This uncertainty, designated H, is the entropy of the information about the particular situation where

$$H(p_1, p_2, \ldots, p_n) = -\sum_{k=1}^{n} p_k \log p_k$$

where $p_1, p_2, \ldots, p_n$ are the probabilities of mutually exclusive events and the logarithms are taken to an arbitrary but fixed base and $p_k \log p_k$ always equals zero if $p_k = 0$. For example, if $p_1 = 1$ and all other ps are zero, the situation is completely predictable beforehand; there is no uncertainty and so the entropy is zero. In all other cases the entropy is positive. *See* INFORMATION THEORY.

In introducing entropy of an information space, C. E. Shannon described a source of information by its entropy H in bits per symbol. The ratio of the entropy of a source to the maximum rate of signaling that it could achieve with the same symbols is its relative entropy. One minus relative entropy is the redundancy of the source. [F.H.R.]

Bibliography: J. G. Aston and J. J. Fritz, *Thermodynamics and Statistical Thermodynamics*, 1959; K. G. Denbigh, *Principles of Chemical Equilibrium*, 1955; A. I. Khinchin, *Mathematical Foundations of Information Theory*, 1957; G. N. Lewis and M. Randall, *Thermodynamics and the Free*

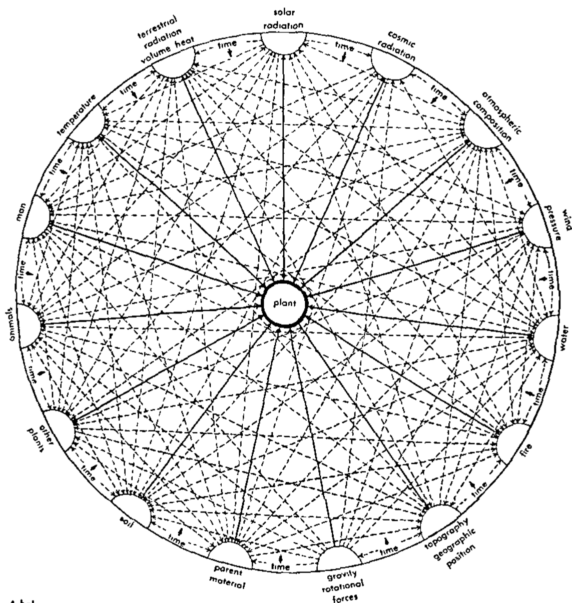
Energy of Chemical Substances, 1923; C. E. Shannon, A mathematical theory of communication, *Bell System Tech. J.*, 27(3):379–423, 27(4):623–656, 1948; R. C. Tolman, *Statistical Mechanics, with Application to Physics and Chemistry*, Am. Chem. Soc. Monograph, 1927.

Environment

Ecologically, the environment is the sum of all external conditions and influences affecting the life and development of organisms. Various ecological principles and concepts have been developed in regard to the environment. Two main aspects of the environment are usually considered, the abiotic and biotic. These divisions are artificial in the sense that neither can be separated when organisms are studied. All environmental aspects and their influences on living organisms must be considered together.

Nature of the environment. Holocoenosis, the nature of the action of the environment, pertains to those factors which exist as a vast complex and therefore do not act separately and independently, but as a whole. This principle, which lies at the core of all ecological thinking, is shown in the illustration.

The particular environment occupied by an organism or by groups and communities of organisms is the habitat. Environmental conditions generally exist in an area independently of the occurrence of any species of plant or animal that may depend on them, and area is used to express this concept, for example, in contrasting humid versus dry climates. The whole system of interaction between a



A holocoenotic environmental complex. (From W. D. Billings)

particular organism and its physical and biotic environment is the niche of that organism. The term indicates not only the habitat, or microenvironment, *but also the activity of the organism.* The environment of small areas in contrast to that of larger ones, or of particular organisms in contrast to generalized environments of communities is recognized as the microenvironment. Other terms commonly applied to this concept, but much more restricted in scope, are microclimate and bioclimate. Being relative, the term may signify the environment of a pine stand, or equally well, a lichen *within that stand.*

Dynamic concept. Variations in space and time are responsible for a dynamic concept of the environment. In space, there is continual change from the equator to the poles, and from sea level to outer space or to ocean depths, and from the forest floor to the tree tops. Variations in time may be geological, as climates, carbon dioxide content of the air, and salt content of the ocean, or be concerned with present day cycles and periodicities. The more familiar include day and night, ocean tides, moon phases, and seasonal cycles.

The change or rate of change in an environmental variable is the gradient. This is pertinent, through all units of space and time, to an understanding of the continuous adjustment life must make to its environment.

Action systems provide a convenient way of referring to the interaction between organisms and their complex environment. Ecological action is the effect of the physical environment on organisms; reaction is the reciprocal effect of organisms on the physical environment; and coaction, the interaction of organisms on each other.

Limiting factors. The concept of limiting factors recognizes that any factor of the environmental complex which approaches or exceeds the limits of tolerance of any organism or group of organisms acts as a control on them. This concept is derived from two laws. The law of the minimum, originally given by Justus Liebig in 1840, has been restated by W. P. Taylor as follows: "The growth and functioning of an organism is dependent upon the amount of the essential environmental factor presented to it in minimal quantity during the most critical season of the year, or during the most critical years of a climatic cycle." By adding the concept of an ecological maximum, V. Shelford in 1913, through his law of tolerance shows that all organisms or groups of organisms must live in a range of conditions between the maximum and minimum which then represent the limits of tolerance. A series of terms has come into general use utilizing the prefixes "steno-" meaning narrow, and "eury-" meaning wide, to indicate the relative degree of tolerance, as for example stenothermal, which indicates a narrow tolerance to temperature.

Other factors. Those factors which upset the delicate balance of an ecosystem are termed trigger factors They cause a chain reaction which may not end until drastic changes have occurred. For example, the permanent addition of water to a desert environment may ultimately change the whole ecosystem.

The substitution of elevation for latitude is an example of a compensating factor which allows plants of northern distribution to grow far southward on high mountain ranges. Some factors such as temperature may substitute for one another without apparent effect on the organism, but factors such as day length are different.

Environmental substances such as nutrients are necessary for the growth and development, or nutrition, of organisms. Thus, inorganic substances provide for the nutrition of green plants, which in turn provide the organic nutrients of nongreen plants and animals. The characteristic pathways in which these substances circulate from the environment to organisms and back again are known as biogeochemical cycles. Oxygen, carbon, and nitrogen are the principal gases involved, while more than 40 minerals are obtained from the crust of the earth, among them phosphorus and potassium. The synthesis of protoplasm by living organisms and their decomposition after death form the energy basis for these cycles and provide for a classification of organisms into producers, consumers, and reducers. *See* BIOSPHERE, GEOCHEMISTRY OF.

ABIOTIC ENVIRONMENT

The physical or abiotic environment includes all those physical and nonliving chemical aspects which exert an influence on living organisms. Among these factors are soils, water, and the atmosphere, as well as the influence of energy from various sources.

Energy. Various forms of energy exert an influence on, or modify, the environment in which organisms are found. With minor exceptions, the ecologist is concerned with radiant energy received directly from the sun, infrared radiation or heat, visible radiation, and ionizing radiation. These energies, coming from the same source, are subject to the same modifying factors, the primary ones being daily and seasonal cycles which in turn are modified by latitude, altitude, slope, exposure, atmospheric conditions, and the color, texture, and cover of the ground surface.

Temperature The intensity aspect of heat energy is temperature, and, with moisture and light, it is one of the most familiar of environmental conditions. In nature, temperature ranges from a low of −70°C (−94°F) recorded in Siberia, to highs of 60°C (140°F) in deserts to almost 100°C (212°F) in hot springs. With few exceptions, however, almost all life exists within the relatively narrow range of −17.8°C (0°F) to 45°C (113°F). The specific responses developed by organisms to temperature include metabolic activity, behavior, abundance, and distribution. With the exception of birds and mammals, the body temperature of plant and animal life is determined primarily by the external environment. *See* HIBERNATION; HOMEOSTASIS; HYPOTHERMIA.

Light. The source of light may be the sun, moon, stars, or luminescence, but usually only the first two are of sufficient intensity to influence life. Light varies according to wavelength, intensity, and duration. Sunlight may vary at noon in the open from a few hundred foot candles on overcast days to more than 14,000 foot candles on clear days in southern latitudes In woods, it may vary from a few to 1000–2000 foot candles under the same conditions. Maximum moonlight is about 0.05 of a foot candle. Light penetration in relatively clear water is about 30% at 5 meters (m) and 10% at 10 m. In addition to the well-known uses of light for photosynthesis in plants and vision in more advanced animals, it also affects growth, development, and survival, and through many types of photoperiodicities, affects the behavior, life cycle, and distribution of many plants and animals

Ionizing radiation. Ultraviolet, nuclear, and cosmic radiations have sufficient energy to damage protoplasm on relatively short exposures. The potential significance of ionizing radiation as an environmental factor was not recognized until the mid-1950s when world-wide attention was focused on ionizing radiation received from fallout. The atmosphere is opaque to all ultraviolet light shorter than 2900 A so that little or no injury results at moderate altitudes In the stratosphere and above, the lethal action of ultraviolet light is important, the bacterial effects beginning at 3650 A units. Nuclear radiation, composed of alpha and beta particles and gamma rays, comes from radioactive substances naturally occurring in the crust of the earth, and from living and nonliving materials which have these incorporated in them. The intensity of this background radiation varies on an average from 0 1 to 0.5 roentgen units per year, although many areas give much higher readings. Due to fallout, this natural radiation is being added to each year The addition in 1956 varied on an average from 0.01 to 0 05 roentgen units per year, with some areas receiving much higher doses. Cosmic rays come from space and vary appreciably with altitude. Some geneticists estimate that 10% of the mutations in geological time are attributable to these ionizing radiations. Very little is yet known about their direct effects at natural levels on organisms.

Water. Water and air are the fundamental media in which life exists. As such, they provide the basis for the division of the world into two major envi- [illegible]

high surface tension, high ionization, high freezing point, latent heat of fusion, latent heat of evaporation, and high specific heat. As the chief components of protoplasm, water and air have many common physicochemical characteristics.

The distribution of water through the atmosphere and over the terrestrial *environment* is controlled by the hydrological cycle. Principal marine environments are tidal zones (neritic), and oceanic zones (euphotic, pelagic, and benthic). Fresh-water environments are standing water (lakes, ponds, swamps, and bogs), and running water (springs, streams, and rivers). On land, the term xeric refers to relatively dry, hydric to relatively wet, and mesic to moist environments. *See* HYDROLOGY.

Atmospheric moisture may exist in three states, the solid (hail, sleet, or snow), liquid (rain), and gas (water vapor). The water vapor content of the atmosphere varies with temperature and is referred to as humidity. The ratio of precipitation to evaporation is the central factor in the distribution of vegetation zones, as desert, grassland, woodland, and forest

Soil moisture varies with the structure and texture of the soil, as well as with precipitation-evaporation factors. Soil is saturated when all pore spaces are filled. The pore spaces make up 40–60% of the volume of soil. Water which subsequently drains away is gravitational water, while that retained by capillary forces is capillary water. In very dry soil, the extremely thin film which cannot move as a liquid is hygroscopic water, and that chemically bound with soil materials is combined water. All gravitational and most capillary water is available to plants, while bound water and hygroscopic water cannot be utilized. Plants which can no longer obtain moisture from the soil reach a physiological state at which wilting occurs. This varies among different plant species and is known as the wilting coefficient of the soil.

Atmosphere. The atmosphere is composed of 78% nitrogen, 21% oxygen, 0.03% carbon dioxide, several other gases in very small quantities, and varying amounts of water vapor. All exist in a simple physical mixture. In addition to its physical characteristics as a medium, the oxygen and carbon dioxide of the atmosphere chemically affect all forms of life through photosynthesis and respiration, that is, through their reciprocal relationship of synthesis and decomposition. Concentration of these gases varies with organic and industrial activity, with altitude in the atmosphere, with organic activity, and with their solubility in fresh and salt water. Carbon dioxide also influences the hydrogen ion concentration of soil and water.

Wind is a significant atmospheric phenomenon acting through its physical force. It exerts an influence on the rate of transpiration in plants and evaporation of moisture. It modifies weather and climate, and is a disseminating agent for pollen, spores, and microscopic organisms.

Substratum. This is any solid on whose surface an organism rests or moves, or within which it lives, thus providing anchorage, shelter, and food. Aquatic substrata are rock, sand, and mud, while terrestrial substrata are rock, sand, and soil, soil being the most complex.

Soil is the link between the mineral core of the earth and life on its surface. It consists of decomposed mantle rock and organic matter, with spaces for water and air, and is organized into distinct lay-

ers or horizons. The A horizon, or surface soil, is the zone of most abundant life, and also of leaching. The B horizon, as subsoil, is a zone of deposition from the layer above, while the C horizon is composed of the parent materials. The major factors in soil formation are climate, topography, living things, and time. The differential nature of these factors is responsible for the great soil groups of the world. *See* SOIL.

Other abiotic factors. These include such conditions as gravity, pressure, sound, and fire. Gravity affects the environment through its action in isostasy, earthquakes, and stratification in air and water. Its effects on organisms are through polarity, orientation, structure, distribution, and behavior. Pressure acts mechanically and environmental pressures range from a half atmosphere pressure at 5800 m elevation to 1000 atmospheres at a depth of 10,000 m in the ocean. Since pressure changes much more rapidly in water than in air, it becomes of paramount importance in the ocean through its effects on the structure, physiology, and distribution of organisms.

Sounds are produced and conveyed by mechanical vibrations, substratal vibrations, and mechanical shock which are interconnected phenomena. While aerial, and usually audible, vibrations are important to man and many animals, particularly birds and mammals, the substratal vibrations such as earthquakes, especially when sudden and intense, are the ones of prime importance in ecology.

Fire has always been an important factor in the terrestrial environment. Classified as crown, surface, and ground types, fires exert direct effects through injury and destruction and indirect effects through often drastic alterations of environmental conditions, and changes in community relationships. The beneficial values of controlled burning in forestry and game management are well established, and fire is an important tool in shifting agriculture in the humid tropics.

BIOTIC ENVIRONMENT

The biotic environment consists of living organisms, which both interact with each other and are inseparably interrelated with their abiotic environment.

Interactions between organisms. Within a population, interactions include such aspects as density, birth rate, death rate, age distribution, dispersion, and growth forms. *See* ECOLOGIC INTERACTIONS.

Interactions between organisms at the interspecies population level include competition, which arises from utilization of the same thing when it is in short supply. Predation and parasitism are negative interactions which result in harm or destruction of a prey or host population by the predator or parasite population. Positive interactions occur such as commensalism in which one population is benefited, cooperation in which both populations are benefited, and mutualism in which both populations are benefited and completely dependent on each other. At the community levels, interactions include dominance, where one or several species control the habitat. In succession, there is an orderly process of community change. Rhythmic change in community activity is termed periodicity.

Interactions with the abiotic environment. These interactions include (1) the effects of an organism on the microenvironment through temperature, water, wind, and light, (2) their modification of the substrate as through soil building, and (3) their modification of the medium, as in aquatic habitats [R.B.P.]

Bibliography: W. C. Allee, A. E. Emerson, O. Park, T. Park, K. P. Schmidt, *Principles of Animal Ecology*, 1949; G. L. Clarke, *Elements of Ecology*, 1954; R F. Daubenmire, *Plants and Environments*, 1947; R. Geiger, *The Climate Near the Ground*, 1950; H. G. Lundegardh, *Environment and Plant Development*, 1931; E. P. Odum, *Fundamentals of Ecology*, 1953.

Environmental protection

The protection of man and equipment against stresses of climate and other elements of the environment. Research in this field aims to improve the effectiveness of human performance under all environmental conditions through interrelated studies of environment, man, and materials, including clothing, shelter, food, machines, and equipment. Systematic research in this field was begun in 1942 to meet the need of the Quartermaster Corps to provide adequate clothing for military personnel overseas in World War II. The principles evolved are equally applicable to civilians under conditions of field or outdoor life. *See* ECOLOGY; ENVIRONMENT.

Physical geographic environment. In order to know the environmental stresses from which men and materiel must be protected, the scientist must study all stressful aspects of the geographic environment, especially climate. Both the essential character and the regional and seasonal distribution of environments require attention. Heat and cold have received primary emphasis, and ambient temperatures have been mapped for all parts of the world. Since 1943, U.S. Army clothing authorization has been based upon clothing zones developed from mean monthly temperature maps. Clothing almanacs show the month-by-month clothing needs. Refinement of the basis of mapping clothing- or equipment-use zones proceeds as data are processed for maximum and minimum temperatures and, most realistic of all, for frequency and duration of critical temperatures. Direct solar radiation and wind also modify the effects of heat and cold induced by basic ambient temperatures. The term windchill is used to express the combined cooling effect of wind and temperature upon the body.

Factors of moisture, including rainfall, snow, glaze, hail, humidity, dew, and fog, introduce other stresses, only in part thermal. Wind is a po mechanical factor that must be considered

design of tents and other structures, and air-borne sand and dust are irritants and abrasives. For work with combined climatic stresses, individual elements are grouped into climatic types. Army clothing, for example, is classified for research purposes into four moisture-temperature types: cold-dry, cold-wet, hot-dry, and hot-wet.

Microclimates are studied to gain an understanding of the conditions experienced by man and materiel at various levels at, above, and below ground, and under diverse terrain or vegetation situations. The designer of shoes, for example, needs to know not merely that the "standard" air temperature in a desert may reach 125°F but also that the ground temperature may reach 160°F.

The terrain, both physical and biological, introduces additional stresses. The composition of the ground surface, such as sand, rough rock, or clay (hard when dry, mud when wet), affects movement of vehicles and adequacy and longevity of clothing. Surface configuration, especially slope steepness, is of paramount importance for vehicular movement, and ways of making quantitative regional generalizations must be devised. Water, in the form of streams, lakes, swamps, or ice, adds other problems of personnel protection. For example, the invention of crevasse detectors has greatly increased the safety of travel on glaciers.

Biological elements of the environment cannot be overlooked. Natural vegetation such as taiga, tropical forest, brush, grassland, desert, and tundra, and cultivated vegetation such as cane fields, rice paddies, and orchards affect durability of clothing, trafficability of vehicles, food and fuel for survival, and concealment against hostile eyes. The camouflager needs to know the seasonal changes of color of vegetation, ground, and snow. Noxious insects are usually the most serious animal pests. Character, seasonal occurrence, and distribution must be studied to permit development of insect control or individual protection with netting or repellents. Microorganisms and air conditions favoring them are part of the problem of minimizing deterioration of food and materiel in storage.

Man. In the environmental protection complex, man himself is studied from several viewpoints. Physiological studies are made under controlled conditions of environment in climatic chambers, or outdoors where all environmental factors converge upon the test subject in the complexity of their interactions. Attention is paid to effects on body and skin temperatures of exposure to heat or cold for varying periods of time and types of activity in which the subject wears various types of clothing. Sweating rates and water requirements, work capabilities, metabolism, and food requirements in extreme environments are observed and measured. To anthropologists falls the task of making body measurements essential for the design of clothing that fits properly to give maximum protection and comfort. Psychologists design questionnaires to elicit statistically meaningful reactions to clothing in varying environments, determine problems of human factors in the use of materiel, study effects of environmental extremes on the psychophysiological capabilities of man, check equipment to assure operability by men clothed protectively, and determine acceptability of protective items to be sure they will be used by the men who need them.

Research on materials. A third element of the environmental complex, materials, is studied to determine its effectiveness in interposing a protective barrier between man and the environment.

Clothing. Studies must deal with the characteristics of textiles such as wind perviousness, insulating value, wicking capability, durability, and comfort of individual garments and of clothing ensembles, through the activity of the biophysicist and the clothing specialist. The widely used "layer system" takes advantage of the insulating value of air layers between cloth layers. Even in arctic clothing, provisions must be made for ventilation to prevent overheating and moisture accumulation. Considerable attention is being paid now to the possibility of developing disposable paper garments to obviate the necessity for laundering. Laundering has its own problems, especially in desert areas where water is limited and in cold areas in which much fuel must be expended to melt ice. Handwear must retain dexterity and provide insulation; headgear must be designed to permit



Fig. 1. Environmental protection testing of cold-weather clothing at temperature of −40°F with gentle breeze, in arctic wind tunnel, Quartermaster environmental research center. Skin temperatures of the men on the treadmill are being measured by means of thermocouples attached to various parts of the body. (U.S. Army photograph)



Fig. 2. Metabolism test of men walking in loose sand in desert heat. (*U.S. Army photograph*)

sight, hearing, and breathing while providing insulation or shade. Trench foot, the dreaded scourge of cold, wet weather, has been nearly eliminated by the development of the insulated boot, with a layer of felt or sponge rubber sandwiched between layers of waterproof rubber—an item invented in the Quartermaster Corps but since applied widely to civilian use. Sleeping gear involves problems associated with the ground surface as well as with the ambient air.

Shelter. Shelter must be designed to protect against extremes of cold, heat, and rain. Unlike clothing, the shelter must be heated mechanically by means of a stove in cold climates, unless it is to be used for the storage of nonperishable articles, thus adding to the great bulk of petroleum fuel products that must be supplied any large modern force of men in the field. In summer in hot climates, ways must be found to reduce the heat load occasioned by intense solar radiation and air temperature. The careful study of geographical sites is an added environmental requirement in the establishment of shelter, especially for permanent and semipermanent buildings.

Relations with materiel testing. The term environmental protection, originally applied to the protection of man against the elements, has been extended to include protection of materials and the development of environment-resistant materials. The whole field of testing paint, vehicles, or electronic equipment against environmental stresses in the field or in the laboratory (including climatic chambers large enough to contain an assembled aircraft) is in a sense a part of environmental protection research. Materiel testing requires that quantitative environmental criteria for design must be formulated in advance, and that the role of man as operator in the system must be considered. In modern materiel testing, much attention is paid to vibration, shock, and acceleration, which are considered to be environmental factors though they are usually man-made rather than natural factors.

Special environmental complexes. Man-made regional features and man's frontiers in space add complexities to environment. The environments resulting from military operations give rise to studies of such matters as the use of body armor against projectiles, physiological and psychological effects of explosions, minimization of radiation effects, and the psychological effects of limited rations. Large solar furnaces have been constructed to obtain radiation intensities that are comparable to those obtained by the use of nuclear reactors.

Recently human environments have been included among the factors to be considered in environmental protection. The social, legal, political, and economic practices and attitudes of local populations have a profound bearing upon the degree of success likely to be achieved by visitors or invaders. Certainly the material portions of man-made environments, the settlements, roads, and forms of land use, may have as direct an influence as the natural terrain and vegetation.

Space environments pose special problems of environmental protection, mostly intensified forms of the ordinary stresses of radiation, vibration, acceleration, and high-altitude conditions, though weightlessness is a space factor without earthly parallel. To the increasing number of all-weather climatic testing chambers, there are now being added chambers to simulate the rarefied conditions and intense radiation of 400 miles and more above the earth. Research on selenography, or lunar "geography," is essential in the advance planning of environmental protection of man on the moon. See CLIMATOLOGY. [P.M.E.]

Bibliography: D. H. K. Lee and H. Lemons, Clothing for global man, *Geograph. Rev.*, 39:181–213, 1949; L. H. Newburg (ed.), *Physiology of Heat Regulation and the Science of Clothing*, 1949; U.S. Army, *Operation of Material under Extreme Conditions of Environment*, AR 705–15, 1957; U.S. Dept. of Defense, *Climatic Extremes for Military Equipment*, MIL-STD-210A, 1957.

Environmental test

Laboratory tests conducted to determine the functional performance of a component or system under conditions that simulate the real environment in which the component or system is expected to operate. The performance of all components and systems is influenced to varying degrees by the environment. Moral, economic, and design considerations often preclude testing for environmental effects by actual operation under real conditions. A man must not be rocketed into space until laboratory testing indicates excellent chances for survival; in toto testing of complex missiles be prohibitively expensive; and, finally, ob

tion under controlled laboratory testing provides invaluable design information.

It is difficult, if not impossible, in a system of any magnitude to define completely the functional requirements to which the system will be subjected. It is equally difficult to define for complex systems those possible combinations of difficult environments which might influence performance most seriously. Usually a series of environmental extremes to which the component may be exposed is defined, and satisfactory functional performance is required during exposure to one or more combinations of these environmental extremes.

Functional failure of a component may be due to deterioration during prolonged storage or transit under difficult environmental conditions. Laboratory environmental testing also explores the resistance of components to such unfavorable conditions.

Temperature tests. Temperature testing in the laboratory is conducted in a sealed chamber, which can be cooled by means of artificial refrigeration to very low temperatures or heated to very high temperatures by means of electrical or combustion heaters. The range of temperatures depends, of course, on the subsequent application of the component, the typical temperature range for military components is −60° to +165°F. Astronautical devices can experience a wider range, from near absolute zero to thousands of degrees F. The test chamber must be capable not only of maintaining a desired temperature for extended periods of time, but also of varying the temperature of the device under test at some predetermined rate, thus simulating the temperature gradients to which the device would be exposed in actual operation or storage, as, for example, heating due to atmospheric friction of the elements in a missile nose cone upon reentry into the earth's atmosphere or the daily temperature cycle experienced by a solid-fuel rocket motor on the surface of the moon.

For instrumentation of the tests, the chamber must have a number of temperature transducers by which local temperatures of various parts of the component under test and chamber air temperatures can be determined and recorded as a function of time. The test chamber should be designed so that the component can actually operate during temperature testing; for example, input-output power [illegible]
gear [illegible]
comp [illegible]
MEASUREMENT.

Humidity, spray, immersion, fungus, and sand tests. Many components and systems must perform satisfactorily during and after exposure to wide ranges of humidity, when subject to fresh- and salt-water spray, or complete immersion, and when blasted by wind-borne sand. Certain combinations of temperature and humidity as experienced in tropical climates are conducive to fungus growth which can deteriorate the performance of components, especially electrical and electronic devices. When specifications require an equipment to operate in such conditions, the environmental testing program must include facilities to duplicate insofar as is possible the character and intensity of conditions to be experienced. Humidity testing is usually done in conjunction with temperature testing. The test chamber is provided with equipment by which the water content of the air can be varied over the desired limits. The water content of the air is measured by either wet-bulb or hair hygrometers together with appropriate recording equipment. *See* MOISTURE-CONTENT MEASUREMENT.

In the case of spray, immersion, and sand testing, specific test conditions must be determined. Military specifications have established standard conditions for such tests.

Pressure tests. While components are normally expected to operate only at average sea-level pressures, other applications may require satisfactory operation from sea-level pressure to low pressures corresponding to high altitudes or outer space, where pressure is so low that it cannot be measured except on a statistical basis. Still other applications require satisfactory performance at high pressures characteristic of great depths beneath the surface of the sea or perhaps in high-pressure transients, such as in the shock waves from high-explosive or atomic blasts. When such requirements are posed, a satisfactory environmental test procedure must provide laboratory facilities for duplicating, in as far as is possible, the actual pressure levels and pressure gradients to which the components will be subjected. Frequently the same chamber in which humidity-temperature testing is conducted is connected to both pressurizing and vacuum pumps in order that the performance of the system may be monitored while the external pressure is varied between prescribed levels and at prescribed rates of change. Appropriate pressure transducers transmit the desired pressure information to recorders which provide a permanent record of the pressures to which the components were subjected while in operation. *See* PRESSURE MEASUREMENT.

Acceleration, shock, and vibration tests. Many components and systems are subject in actual operation to high accelerations in one or more orthogonal directions, vibrations of various amplitudes and frequencies, and shocks of various intensities. For instance, the control components in a missile-rocket-engine orienting system are subject to high accelerations during launch, high decelerations during atmospheric reentry, shocks (high time rates of change of acceleration) during missile and booster separation at booster burn-out, high-frequency vibration from rocket-engine-generated noise and high-amplitude, low-frequency vibrations due to missile-air-frame elasticity. In order to provide some assurance of the satisfactory performance of such a component in the actual flight test of a missile, it is essential to duplicate, in the laboratory, with all appropriate accuracy, as many as possible of these several environmental

factors, and to explore the performance of the component while it is subjected to one or more of these disturbing influences.

Acceleration tests. Acceleration testing is conducted on a centrifuge or a sled. In a centrifuge the component is mounted at the end of an arm subject to high rotative speeds. The resulting centrifugal acceleration on the component can be adjusted by varying the speed. Accelerations of one hundred times that of gravity with fairly massive devices are readily achievable, but rate of change of acceleration is limited. Component operation and instrumentation connections are brought off the centrifuge by means of rotating connections at the center of rotation of the arm. Since the actual application may expose the component to accelerations in more than one direction, the component is successively oriented and tested in several positions with respect to the arm. *See* CENTRIFUGATION.

Acceleration testing in pure translation is accomplished by means of rocket-accelerated sleds on which the components to be tested are mounted. The sleds are guided by railroad rails laid exceptionally straight and flat for distances of several miles. Acceleration tests conducted on sleds can closely simulate the acceleration-rate program to which the component will be exposed in actual missile flight. The sleds are decelerated by means of a water brake—a scoop on the sled is forced to travel through a trough of water—thus reducing the velocity of the sled very rapidly. By such means the actual deceleration which a missile component might experience, for example, on atmospheric reentry, can be duplicated. When sled testing is conducted with the component operative, the apparatus to record the salient performance parameters of the component is carried on the sled. *See* ROCKET-SLED TESTING.

Shock tests. Shock testing consists of exposing the component to high rates of change of deceleration or acceleration, by either dropping the component onto a resilient or absorbent surface such as rubber, cork, or sand, or subjecting it to a pressure pulse from a compressed gas. In the latter case the component is mounted on a structure connected to a piston in a cylinder. High-pressure compressed gas is rapidly admitted to the cylinder, usually by the prearranged failure of a diaphragm between the cylinder and the pressure source.

Shock apparatus is usually experimentally calibrated in terms of the change in gravitational acceleration units per millisecond occurring in the deceleration or acceleration; a rate of 100 times gravity per millisecond is typical. In the drop test, the shock intensity is changed by either varying the elevation through which the object falls or varying the resilience of the absorbing media. For pneumatic shock testers, the intensity can be altered by changing the initial pressure of the gas admitted to the cylinder.

Vibration tests. The ability of a component to perform satisfactorily while subjected to vibration is evaluated by mounting the component on a structure whose frequency and amplitude of vibration can be controlled. Vibrators, or shakers, are driven mechanically or electromechanically. In mechanically driven vibrators the table is connected by an eccentric mechanical linkage to an electrical motor, or the resiliently mounted test table also supports an electric motor whose rotor shaft has an intentional unbalance. The frequency and amplitude of vibration may be altered, for the former, by adjusting the link eccentricity and speed and for the latter by changing the motor speed and magnitude of unbalance. Typical ranges for mechanically driven tables are between 3 and 500 cps at amplitudes which correspond to maximum accelerations of up to 8 times that of gravity.

Electromechanical vibrators are essentially large loudspeakers. The component to be tested is connected to the vibrating cone. When an electric signal of the desired frequency and amplitude is impressed upon the electromagnetic driver of the cone, the component is made to vibrate at the desired frequency and amplitude.

Vibration-test programming takes either of two general forms The first consists of exposing the part, in time sequence, to sinusoidal displacements over a monatomic range of frequencies and a corresponding range of amplitudes, for instance, a frequency sweep from 3 to 500 cps with a maximum amplitude of eight times gravity, each frequency sweep taking 10 sec. When vibration resistance in more than one orthogonal direction is required, the component is tested while mounted in two or more positions.

The alternative programming technique for components, which is more typical of the actual service conditions, consists of exposing the component to a random vibration program which simulates the ranges and amplitudes of vibrations which can be anticipated in actual service. The random vibration program is either assembled mathematically or is recorded in an actual service test and played back to drive the electromechanical vibrator in the laboratory. *See* SYSTEMS ENGINEERING; *see also* AIRCRAFT TESTING. [R.W.M.]

Enzyme

A catalyst produced by living cells. Most of the chemical reactions that take place in living organisms are catalyzed by specific enzymes, without which such reactions would proceed at a negligible rate, if at all. Living organisms are capable of carrying out hundreds of different metabolic reactions and therefore possess a comparable number of individual enzymes. Some representative enzymes are shown in the table.

Nature of enzymes. Many enzymes have been purified, and some have been isolated in crystalline and homogeneous form. All enzymes thus far investigated have been found to be proteins. Thus, they are denatured by heat, they can usually be precipitated from aqueous solutions by ethanol, trichloroacetic acid, or by high concentrations of inorganic salts, and they cannot pass through

Some representative enzymes

Enzyme	Some sources	Reaction catalyzed
Pepsin	Gastric juice	Hydrolysis of proteins to peptides and amino acids
Urease	Jack bean, bacteria	Hydrolysis of urea to ammonia and carbon dioxide
Amylase	Saliva, pancreatic juice	Hydrolysis of starch to maltose
Phosphorylase	Muscle, liver, plants	Reversible phosphorolysis of starch or glycogen to glucose-1-phosphate
Transaminase	Many animal and plant tissues	Transfer of an amino group from an amino acid to a keto acid
Phosphohexose, isomerase	Muscle, yeast	Interconversion of glucose-6-phosphate and fructose-6-phosphate
Pyruvic carboxylase	Yeast, bacteria, plants	Decarboxylation of pyruvate to acetaldehyde
Catalase	Erythrocytes, liver	Decomposition of hydrogen peroxide to oxygen and water
Alcohol dehydrogenase	Liver	Oxidation of ethanol to acetaldehyde
Xanthine oxidase	Milk, liver	Oxidation of xanthine and hypoxanthine to uric acid

permeable membranes. Some enzymes are simple proteins; that is, they yield only amino acids on hydrolysis. Others are conjugated proteins; these contain a nonprotein component, bound to the protein, called a prosthetic group. If separated from each other, neither the protein component (the apoenzyme) nor the prosthetic group (or coenzyme) alone possesses the full catalytic properties of the complete compound (the holoenzyme). However, the initial activity may often be restored by recombining the apoenzyme and the coenzyme.

Enzymes measurably accelerate chemical reactions, even in extremely low concentrations. An enzyme-catalyzed reaction, like any other chemical reaction, proceeds spontaneously, at constant temperature and pressure, only if accompanied by a decrease in free energy; the equilibrium concentrations of reactants and products are the same in the presence of a catalytic amount of enzyme as in its absence. A given amount of enzyme can catalyze an indefinite amount of chemical change without itself being diminished. However, since most isolated enzymes are relatively unstable, they usually lose activity under the conditions employed for their study. Unlike many catalysts used in organic chemistry, enzymes do not require high temperatures to exert their effect, but act optimally in the range of 20–40°C. High temperatures, on the contrary, rapidly inactivate and denature enzymes. Most enzymes lose their catalytic activity after being heated at 100°C for even a few minutes.

Since enzymes are large molecules, and the substances upon which they act, that is, their substrates, are often of relatively low molecular weight, it appears possible that the entire protein molecule may not be essential for enzymatic activity. Catalase, for example, has a molecular weight of approximately 250,000, while its substrate, hydrogen peroxide, has a molecular weight of 34. There is some evidence that only a portion of the protein molecule, the active center, is responsible for its catalytic properties. Thus, certain enzymes have been partially degraded by enzymatic hydrolysis without loss of activity. For example, the proteolytic enzyme papain has been reported to be fully active after approximately two-thirds of its 180 amino acids have been removed from one end of the peptide chain. The fact that all of the free amino groups of pepsin may be acetylated without loss of activity indicates that these groups are not necessary for the catalytic action of this enzyme. Two other proteinases, trypsin and chymotrypsin, have been found to react stoichiometrically with diisopropylfluorophosphate with complete loss of activity. It has been postulated that this inhibitor reacts with an essential group in the active center of the enzyme. See ACETYLATION; STOICHIOMETRY.

Some enzymes are produced by cells in the form of an inactive precursor, called a zymogen. Thus, pepsin is secreted by the gastric mucosa as an inactive protein, pepsinogen; this is converted to the active enzyme by the hydrochloric acid present in the stomach, as well as by pepsin itself. Similarly, chymotrypsinogen is converted to chymotrypsin by trypsin, an activation which has been shown to involve the hydrolytic cleavage of a specific peptide from the zymogen.

Isolated enzymes may be inactive because of the removal of an essential cofactor during purification; addition of the cofactor restores enzymatic activity. The cofactor may be an organic coenzyme, or may be a specific metal ion, such as Mg^{++}, Mn^{++}, Co^{++}, or Ca^{++} (magnesium, manganese, cobalt, calcium). Other metals that have been found to be an integral part of certain enzymes and essential for their activity are iron, zinc, copper, and molybdenum. Metalloenzymes are, in general, strongly inhibited by chelating agents, or by substances which form complexes with polyvalent metals, such as iron, copper, and magnesium. Enzymes that contain sulfhydryl groups essential for activity may be activated by the addition of cysteine or glutathione and inhibited by reagents that combine specifically with sulfhydryl groups, such as iodoacetate and p-chloromercuribenzoate. Some enzymes may also be inhibited by structural analogs of their substrates. See CHELATION; ENZYME INHIBITION.

Nomenclature. A systematic nomenclature for enzymes has not yet been adopted. Many enzymes

are named by adding the suffix "ase" to the name of their substrates or to the name of the reaction that they catalyze. Thus, the enzyme that catalyzes the oxidation of uric acid is called uricase, enzymes that catalyze hydrolysis of proteins are called proteinases, and the enzymes that catalyze the transfer of amino groups from one molecule to another are known as transaminases. Enzymes isolated from different sources may catalyze the same reaction. However, such enzymes are not necessarily identical and therefore the name of an enzyme should also include its source, as in rat liver phosphorylase. Many enzymes retain the unsystematic name originally given to them; examples are the proteolytic enzymes, pepsin and trypsin, and a yellow enzyme isolated from yeast by O. Warburg in 1932 (the first yellow enzyme to be discovered), which is known as the old yellow enzyme. *See* YEAST.

Purification. Since most enzymes are relatively unstable, special methods are needed for their isolation. The enzyme is first extracted from the cells or tissue in which it is found. This is often accomplished by stirring the ground-up tissue in dilute buffer or in water. Enzymes are extracted from free cells and microorganisms by repeated freezing and thawing, by grinding with abrasives, by exposure to ultrasonic vibrations, and by other methods. Although many enzymes can be made soluble in this way, some remain associated with insoluble particles and may be difficult to obtain in solution. The desired enzyme is next isolated from the complex mixture of proteins and other substances present in the crude extract. Only some of the methods used will be mentioned, because no generally applicable purification procedure exists. The most satisfactory method for each enzyme must be determined empirically. One method widely employed for enzyme purification is ammonium sulfate fractionation. Some proteins are precipitated at relatively low concentrations of this salt, while others remain in solution even in nearly saturated ammonium sulfate. Thus, by increasing the salt concentration in a stepwise fashion and by isolating each precipitate, fractions of considerable purity may be obtained. Salts and organic impurities of low molecular weight are removed from enzymes by dialysis; the enzyme solution is placed in a bag made of a semipermeable material, such as collodion or cellophane, and is immersed in water or in a buffer solution. The protein remains in the bag, while the solutes of low molecular weight diffuse out. Enzyme purification is usually carried out in the cold, at 0–10°C, to minimize losses due to denaturation. However, certain enzymes are stabilized by their substrate, and under these conditions they may be heated for a short time, without loss of enzymatic activity, to a temperature at which other proteins are denatured and precipitated.

Enzymes contain ionizable groups, such as carboxyl groups, which may be negatively charged, and amino groups, which may be positively charged. Since the degree of ionization of each group depends on the pH of the solution, the total charge on the protein varies greatly with the pH. The pH at which the net charge of the protein is zero is called its isoelectric point, and this is usually the pH at which solubility is minimal. Thus, an enzyme may often be selectively precipitated from a solution by bringing the pH to its isoelectric point. Enzymes may also be purified by electrophoresis, which involves the application of an electric field to the solution. The negatively charged proteins migrate toward the anode, at a rate determined by the magnitude of each charge, while the positively charged proteins move toward the cathode; isoelectric proteins do not migrate. Other methods frequently used include precipitation at low temperatures with organic solvents, such as acetone or alcohol; selective adsorption on gels, such as calcium phosphate and aluminum hydroxide, followed by selective elution of the enzyme; and chromatography, on columns of ion-exchange resins or modified cellulose. The amount of enzyme present in each fraction is assayed by measuring the velocity of the reaction catalyzed by a given amount of the enzyme solution under strictly standardized conditions. *See* CHROMATOGRAPHY; ELECTROPHORESIS.

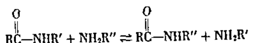
Certain enzymes may also be purified by crystallization. J. B. Sumner prepared the first crystalline enzyme, urease, in 1926, and since then a number of other enzymes have been crystallized (Fig. 1). Crystallinity is not proof of absolute purity; enzyme preparations may contain several components, even after repeated crystallization. The most reliable evidence for the absolute purity of a protein is obtained by physical methods, such as sedimentation in the ultracentrifuge, electrophoresis at several different values of pH, chromatography, and solubility measurements. *See* ULTRACENTRIFUGE

Specificity. The high degree of specificity exhibited by enzymes is one of their most characteristic properties. The majority of enzymes exhibit absolute reaction specificity, that is, they catalyze only one type of reaction. Some enzymes, however,



Fig. 1. Crystalline enzymes. (a) Crystals of beef blood catalase. (b) Crystals of urease from jack bean meal.

are able to catalyze more than one type of reaction. For example, certain peptidases catalyze both hydrolysis of peptide bonds and transamidation reactions; in the latter an acyl group (C=O) is transferred from the amino group (NH_2) of one molecule to that of another:

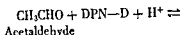


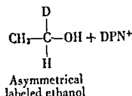
derivatives are unaffected by this enzyme.

Other enzymes may act on a number of compounds that possess a particular chemical group. For example, alkaline phosphatase hydrolyzes a variety of phosphate esters, such as phenylphosphate, glycerophosphate, and others. This type of specificity may be designated as group specificity.

Many enzymes act only on one optical isomer of a compound, thus exhibiting stereochemical specificity (*see* STEREOCHEMISTRY). Conversely, when a compound containing an asymmetric group (for instance a carbon atom with a different group on each of its four valences) is enzymatically synthesized from an optically inactive precursor, only one of the possible optical isomers is generally formed An example of a group-specific as well as stereospecific enzyme is L-amino acid oxidase. This enzyme catalyzes the oxidation of a variety of compounds of the type R—CH(NH_2)COOH, although the rate of oxidation varies greatly, depending on the nature of the R group. No reaction takes place when a substituent is present on either the carboxyl group (COOH) or the amino group. This enzyme is absolutely specific for amino acids of the L configuration. D-Amino acids, which are unaffected by this enzyme, are oxidized by another group-specific enzyme, D-amino acid oxidase. Glycine, which does not have an asymmetric carbon atom, is unaffected by these enzymes but is oxidized by another enzyme When an amino acid is synthesized by a reversal of this reaction, L-amino acid oxidase forms amino acids only of the L configuration, while D-amino acid oxidase forms D-amino acids exclusively.

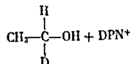
An enzyme may act asymmetrically on a symmetrical substrate to yield a symmetrical product. Such specificity may be detected by the use of isotopes. Thus, alcohol dehydrogenase, which catalyzes the conversion of ethanol to acetaldehyde, removes a specific hydrogen atom from carbon atom 1 of ethanol. This hydrogen atom is transferred to the nicotinamide ring of diphosphopyridine nucleotide (DPN). If reduced DPN containing deuterium, D, is used in the reversal of this reaction, asymmetrically labeled ethanol is produced. No deuterium is found in the acetaldehyde arising from enzymatic dehydrogenation of this ethanol.





When acetaldehyde containing deuterium on carbon atom 1 is hydrogenated with unlabeled DPN, the other isomer of deuterium-labeled ethanol is formed, and on subsequent enzymatic dehydrogenation, all the deuterium is found in the acetaldehyde.





Classification. Enzymes may be classified on the basis of their metabolic function, their chemical structure, their localization within the organism, or the reaction they catalyze. On the basis of reaction specificity, enzymes may be placed in the following categories: (1) hydrolyzing enzymes, or hydrolases; (2) condensing and cleaving enzymes; (3) oxidation-reduction enzymes; (4) isomerizing enzymes, or isomerases; (5) transferring enzymes, or transferases. Several enzymes are difficult to fit into the above categories, while others may fall in two or more groups. Examples of each group are given.

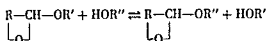
Hydrolases This group includes (1) the digestive enzymes that degrade proteins, fats, and carbohydrates into smaller units which can be absorbed by the intestine; (2) intracellular enzymes that catalyze the cleavage of esters, amides, and acid anhydrides, with the introduction of the elements of water.

Proteinases and peptidases catalyze the hydrolysis of peptide bonds and occur not only in digestive secretions but intracellularly in many tissues; the latter enzymes are designated as cathepsins. Some peptidases catalyze the hydrolysis of esters, as well as transamidation and transpeptidation reactions.

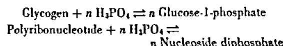
The esterases include the lipases, which catalyze hydrolysis of neutral fats to glycerol and fatty acids; the sulfatases, which hydrolyze esters of sulfuric acid; and the phosphatases, which hydrolyze esters of phosphoric acid. Glucose-6-phosphatase is an example of a phosphomonoesterase. Phosphodiesterases hydrolyze specifically one of the ester bonds of doubly esterified phosphates, such as those present in nucleic acids.

Glycosidases are enzymes that catalyze hydrolysis of glycosidic bonds. An example is α-glucosidase, a group-specific enzyme that also exhibits high stereochemical specificity. It has no effect on β-glucosides or on α-galactosides, which are hydrolyzed by

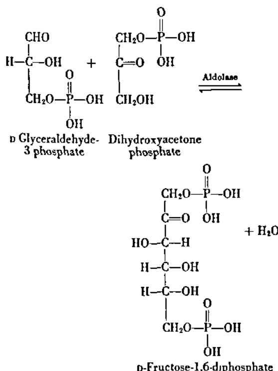
different enzymes. This enzyme also catalyzes transglucosidation reactions:

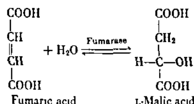


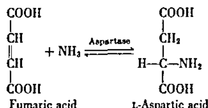
Phosphorylase is an enzyme that catalyzes a reaction analogous to hydrolysis, in which the elements of phosphoric acid are introduced in place of water. Thus, muscle phosphorylase catalyzes the reversible conversion of glycogen to glucose-1-phosphate in the presence of inorganic phosphate. An analogous reaction is catalyzed by polynucleotide phosphorylase:



Condensing and cleaving enzymes A wide variety of enzymes catalyzes condensation and cleavage reactions, which may involve cleavage or synthesis of a carbon to carbon bond, or the removal or addition of a molecule of water, ammonia, carbon dioxide, or other substance. Thus, the reactions catalyzed by aldolase, fumarase, and aspartase are the following:





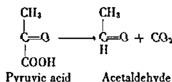


The hydration of carbon dioxide to carbonic acid



is catalyzed by carbonic anhydrase, which plays an important role in the transport of carbon dioxide by blood.

Many decarboxylation reactions are known; one of the first to be discovered was the decarboxylation of pyruvic acid by an enzyme obtained from yeast:



This enzyme, named carboxylase, was shown to require both thiamine pyrophosphate, also known as cocarboxylase, and magnesium ions.

Oxidation-reduction enzymes. Most of the enzymatic oxidation reactions are actually dehydrogenations or transfers of electrons from the substrate, the electron donor, to an oxidizing agent, the electron acceptor. Molecular oxygen usually serves as the ultimate electron acceptor in biological oxidations, but many oxidizing enzymes are unable to use oxygen directly, and therefore the hydrogen atoms, or electrons, must be transferred to intermediate compounds. See BIOLOGICAL OXIDATION.

Enzymes that catalyze dehydrogenations but cannot utilize molecular oxygen as the electron acceptor are called anaerobic dehydrogenases Most of these transfer hydrogen, reversibly, either to diphosphopyridine nucleotide (DPN, coenzyme I, cozymase) or to triphosphopyridine nucleotide (TPN, coenzyme II). See DIPHOSPHOPYRIDINE NUCLEOTIDE (DPN); TRIPHOSPHOPYRIDINE NUCLEOTIDE (TPN). Several of these enzymes have been found to contain zinc. While some are absolutely specific for DPN or TPN, others may use either nucleotide.

Enzymes that catalyze oxidation of substrates by molecular oxygen are called aerobic dehydrogenases, or oxidases This group includes the class of flavoproteins, or yellow enzymes. The yellow color is due to the prosthetic group, which may be either flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD). Although most flavoenzymes are specific for either FAD or FMN, others are active with both. Thus, the old yellow enzyme, which catalyzes the oxidation of reduced TPN, is about 70% as active with FAD as with FMN. Hydrogen is transferred by the enzyme from the substrate to the flavin nucleotide, which is colorless

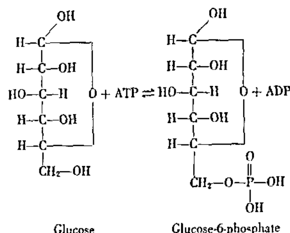
when reduced. It may be subsequently transferred to oxygen or to another acceptor. When oxygen is the acceptor, hydrogen peroxide is produced. Several enzymes of this class have been found to contain a metal and are therefore known as metalloflavoproteins. For example, xanthine oxidase from milk contains iron as well as molybdenum, in addition to FAD.

Another group of oxidation-reduction enzymes contains an iron porphyrin as a prosthetic group. These include the cytochromes and the hydroperoxidases (*see* CYTOCHROME). The cytochromes are important in cellular respiration, since they serve as the final links, in the electron transport chain, between oxidizable substrates and oxygen. A cytochrome may accept an electron from either a flavoprotein or another cytochrome, and may transfer it to either a cytochrome or oxygen. The transfer of electrons from the cytochrome system to oxygen is much more rapid than from flavoproteins to oxygen. Thus cytochromes may serve as efficient intermediates in the reoxidation of flavoproteins. Moreover, instead of hydrogen peroxide, which is a toxic substance, water is produced in the reduction of oxygen by cytochromes. The iron of the prosthetic group passes, reversibly, from the ferrous state in the reduced cytochrome to the ferric state when the enzyme is oxidized.

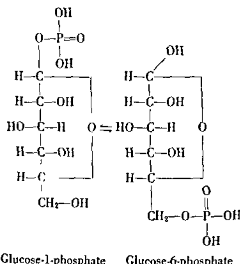
The hydroperoxidases include catalase, which decomposes hydrogen peroxide to oxygen and water, and peroxidase, which catalyzes the oxidation of various substances, for instance, certain phenols, by hydrogen peroxide. Catalase has been found to possess peroxidase activity as well; thus, it can effect the oxidation of ethanol to acetaldehyde in the presence of hydrogen peroxide. Several biological oxidations are catalyzed by copper-containing enzymes. Some oxidation reactions can be shown, by isotope experiments, to involve the direct introduction of molecular oxygen into the substrate. Enzymes catalyzing such reactions have been named oxygenases.

Isomerases. There are several enzymes that catalyze the isomerization of their substrate. Thus, the enzyme phosphohexoisomerase catalyzes the interconversion of glucose-6-phosphate and fructose-6-phosphate. Alanine racemase, which occurs in a number of microorganisms, catalyzes the conversion of either L- or D-alanine to the racemic mixture. Other isomerases, such as those catalyzing cis-trans and keto-enol transformations, are also known.

Transferases. This group of enzymes transfers a group from one molecule to another. The transphosphorylases, which transfer phosphate radicals, constitute an important subclass of this group, since phosphate bonds play a major role in energy transformations of living organisms. Enzymes that catalyze the transfer of the terminal phosphate group of adenosinetriphosphate (ATP) to specific compounds are termed kinases. For example, hexokinase catalyzes the formation of glucose-6-phosphate from glucose and ATP. Mg^{++} is required for the reaction.

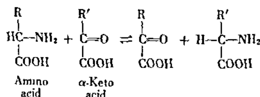


The transfer of a group from one part of a molecule to another may also be enzyme-catalyzed. Phosphoglucomutase, for example, catalyzes the interconversion of glucose-1-phosphate and glucose-6-phosphate.



This reaction requires cysteine and Mg^{++}, as well as catalytic amounts of glucose-1,6-diphosphate. This enzyme may also be classified as an isomerase.

The transaminases represent another important group of transferring enzymes. These enzymes catalyze the transfer of an amino group from one molecule, usually an amino acid, to another, usually an α-keto acid, without the intermediate formation of ammonia.



Transaminases require pyridoxal phosphate or pyri-

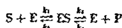
doxamine phosphate as a coenzyme. These coenzymes, as well as a number of others, are derived from vitamins. *See* COENZYME.

Kinetics. Much information may be obtained concerning the mechanism of an enzyme-catalyzed reaction by kinetic studies, that is, studies of reaction rates under various conditions, for example, the reversible reaction



$$S \overset{E}{\rightleftharpoons} P$$

where S is the substrate, P the product, and E the enzyme. Under constant and suitable conditions of temperature and pH, the rate of the reaction depends on the concentration of substrate [S], the concentration of product [P], and the concentration of enzyme [E]. The reaction velocity in the absence of enzyme is negligible. The velocity of the over-all reaction may be obtained by determining the changes in [S] or [P] with time. If a small quantity of enzyme is added to a solution containing only S at an initial concentration $[S]_i$, the velocity of the forward reaction will gradually decrease with time, as [S] decreases, and that of the backward reaction will increase with time, as [P] increases, until eventually equilibrium will be attained and no further net change will be observed. To study the kinetics of the forward reaction alone, one usually determines the velocity over a short time interval at the beginning of the reaction, when [P] is still negligibly small. This has the additional advantage of minimizing the effect of enzyme inactivation during the experiment. For a given value of $[S]_i$, if [E] is much smaller than $[S]_i$, the initial velocity, v_i, is proportional to the enzyme concentration. For a given concentration of enzyme, however, the initial velocity varies as a function of initial substrate concentration (Fig. 2). At low substrate concentrations, v_i is proportional to $[S]_i$ but as $[S]_i$ is increased, v_i approaches a maximum value v_{max} and becomes independent of substrate concentration. A theory explaining these observations was first proposed by L. Michaelis and M. L. Menten in 1913. They postulated that the substrate combined, reversibly, with the enzyme to form an enzyme-substrate complex ES, which in turn decomposed to yield the product and the free enzyme. The latter could then react with more substrate, and the cycle would be repeated:



$$S + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_4}{\overset{k_3}{\rightleftharpoons}} E + P$$

The velocities for these four reactions are expressed, according to the law of mass action, as

$$v_1 = k_1\,[S]\,[E]$$
$$v_2 = k_2\,[ES]$$
$$v_3 = k_3\,[ES]$$
$$v_4 = k_4\,[E]\,[P]$$

where k_1, k_2, k_3, and k_4 are the respective rate constants.

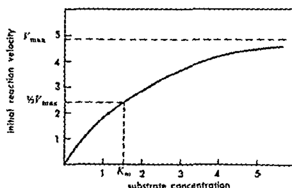


Fig. 2. Initial velocity of an enzyme-catalyzed reaction as a function of substrate concentration.

Michaelis and Menten formulated the Michaelis-Menten equation which relates the velocity of an enzymatic reaction to the concentration of substrate. The Michaelis constant is K_m.

$$[S]_i\left(\frac{v_{max}}{v_i} - 1\right) = \frac{k_2 + k_3}{k_1} = K_m$$

This equation fits remarkably well with the kinetic data obtained from a large number of enzymes. When v_i is equal to $\frac{1}{2}v_{max}$, K_m equals $(S)_i$. Thus, K_m represents the concentration of substrate required to obtain one-half the maximum velocity with a given concentration of enzyme. The Michaelis constant, with respect to a particular substrate, is a characteristic property of each enzyme, and is independent of enzyme concentration. When k_3 is negligible compared to k_2, K_m becomes equal to the dissociation constant of the enzyme-substrate compound, and thus is a measure of the affinity of an enzyme for its substrate.

Although a simple example has been selected here for illustration, most enzymatic reactions involve more than one substrate, and a number of different enzyme-substrate complexes may be formed before the appearance of product and the regeneration of free enzyme. The exact mechanism of a reaction and the number of intermediate complexes involved cannot usually be determined by over-all kinetic studies alone. However, in some cases, direct evidence for the formation of enzyme-substrate complexes has been obtained. Thus, horseradish peroxidase, an enzyme that catalyzes the oxidation of certain compounds by hydrogen peroxide, yields a green complex in the presence of this reagent. Upon addition of an appropriate reducing agent, a red complex is transiently formed before the enzyme finally returns to its original brown color. Some of these changes are extremely rapid, but they can be followed spectrophotometrically by special techniques.

Chemical reactions are usually accelerated by an increase in temperature because molecules require a minimum quantity of energy to react, and at a higher temperature a greater proportion of the

molecules possess this energy. The dependence of the rate constant of a reaction on temperature was expressed by S. Arrhenius in 1889 by the following equation:

$$\frac{d \ln k}{dT} = \frac{E_a}{RT^2}$$

where k is the rate constant, R is the gas constant, T is the absolute temperature, and the proportionality constant E_a is called the energy of activation for the reaction. This may be calculated by measuring the rate of a reaction at different temperatures and substituting the values obtained in the integrated Arrhenius equation:

$$\ln \frac{k_2}{k_1} = \frac{E_a}{R}\left(\frac{1}{T_1} - \frac{1}{T_2}\right)$$

The activation energy is usually much lower for an enzyme-catalyzed reaction than for the same reaction in the absence of enzyme. When the rate of an enzymatic reaction is determined as a function of temperature, it is found that the rate increases with temperature over a certain range, and then decreases sharply because of denaturation of the enzyme as the temperature is increased further. The energy of activation for the denaturation process is usually much higher than that of the reaction catalyzed by the enzyme, and the results obtained

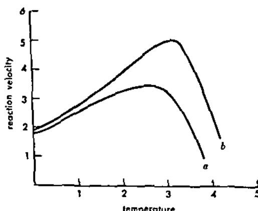


Fig. 3. Effect of temperature on the velocity of an enzymatic reaction. Activity measurements for curve *a* taken over longer time interval than for curve *b*.

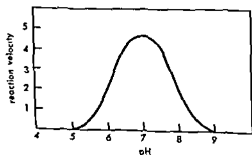


Fig. 4. Velocity of an enzyme-catalyzed reaction as a function of pH.

represent the combination of these two effects. The optimal temperature, as well as the maximum activity observed, depends on the time interval over which the reaction is measured (Fig. 3). Since the stability of the enzyme also varies with pH, substrate concentration, and the presence of impurities, the curves obtained will also depend on these, as well as on other experimental conditions.

Enzyme activity is considerably affected by the hydrogen-ion concentration of the reaction medium (Fig. 4). An enzyme may exist in several different ionic forms, and the ionic species that predominates is a function of pH. Often only one of these forms is catalytically active, or one form may be more active than the others The pH optimum of different enzymes varies over a broad range, from about pH 2 for pepsin to approximately pH 10 for arginase. Some enzymes exhibit a broad optimal range, extending over several pH units, while others have very little activity even one pH unit above or below optimal value.

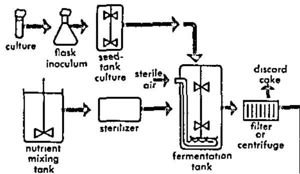
Although the pH optimum is a more characteristic property of an enzyme than its temperature optimum, the pH optimum may also be found to vary with experimental conditions, such as the nature of the substrate, the buffer employed, and the duration of the experiment. Thus, if the substrate contains ionizable groups, the pH optimum for the reaction may depend on the dissociation constants of these groups. The stability of an enzyme at various values of pH is also important, as a decrease in the rate of the catalyzed reaction with changing pH may be due to an increased rate of enzyme denaturation. *See* BACTERIAL ENZYME; BIOCHEMISTRY. [A.ME.; D.WE.]

Bibliography: M. Dixon and E. C. Webb, *Enzymes*, 1958; A. H. Mehler, *Introduction to Enzymology*, 1957; J. H. Northrop, M. Kunitz, and R. M. Herriott, *Crystalline Enzymes*, 2d ed., 1948.

Enzyme (industrial production)

Microbiological processes used industrially to make large quantities of enzymes. Enzymes are organic catalysts, produced by all living cells, which govern the chemical reactions necessary for life processes. Once elaborated by the cell, an enzyme can act independently if proper conditions are maintained (*see* ENZYME). Through the ages, man has made extensive use of enzyme-catalyzed reactions in which the intact cells are utilized for such processes as brewing, baking, and antibiotic synthesis (*see* ANTIBIOTIC; FOOD ENGINEERING). The possibility for separation of the enzyme system from its source has greatly widened the scope of usefulness of enzymes and given rise to the industrial enzyme industry. Although no sales figures are published, it is known that annual sales of enzymes for use in the textile, leather, food and beverage, and pharmaceutical industries total several million dollars.

Industrial enzymes are obtained from three major sources: (1) plant—malt diastase, papain, bro-

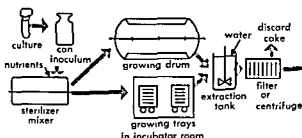


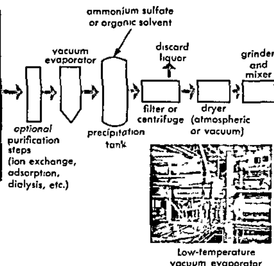


Stainless steel culture cans and shaker flasks



Growing drums





Low-temperature vacuum evaporator

Procedures for enzyme production by industrial fermentation. (*Wallerstein Co., Inc*)

melain, ficin; (2) animal—pancreatic enzymes, pepsin, lipase, catalase, rennin; (3) microbial—amylases, proteases, pectinases, invertase, and glucose oxidase.

The number of enzymes which can be recovered economically from plants and animals is limited. Some are obtained as by-products and thus are limited in availability. Animal enzymes, for example, are supplied by the meat-packing industry and the size of the slaughter limits the total enzyme supply.

The development of fermentation methods for the production of enzymes by microorganisms has assured an unlimited supply. It has also made possible the development of new systems which are not commercially available from plant or animal sources. In the strict biochemical sense, the word fermentation refers to anaerobic transformation of carbohydrates into simpler compounds (*see* BACTERIAL PHYSIOLOGY). Through common usage, this term is now applied to any industrial process involving the chemical reactions of living microorganisms.

Bacteria, fungi, and yeasts are used as enzvme sources. Although details of the processes varv, depending on the organism and system involved, all processes have many elements in common. They fall into two major groups, production by submerged fermentation and production in semisolid culture.

Production by submerged fermentation. The majority of bacterial enzymes and some of the fungal enzymes are produced by submerged fermentation. Stainless steel tanks ranging in size from 1000 to 25,000 gallons are employed. Sterilization is accomplished by use of steam at 121°C for ½–1 hour. Air, sterilized by filtration through activated carbon or glass wool, is supplied through a nozzle or perforated ring at the base of the tank. An agitator is provided to mix the air and the medium. Temperature is controlled by circulating water through a jacket or coil on the tank. Each system has its own optimum temperature, but most fall within the range of 25–37°C. Refrigerated water usually must be circulated to remove the heat produced by microbial action in the tank. Auxiliary tanks are provided for the addition of nutrients during growth, acid or alkali to control pH, antifoaming agents such as lard oil, or silicone compounds.

A pure culture of a selected strain of the microorganism is grown in laboratory shaker flasks. Selection of the proper strain is of extreme importance. These flasks are used to inoculate seed tanks which in turn are used to inoculate production tanks. The organisms used for some of the more important industrial enzymes are listed in Table 1.

Growth media are composed of mixtures of carbohydrates (starch, glucose, molasses, corn syrup), nitrogenous compounds (protein digests, ammonium salts), growth stimulants (yeast hydroly corn steep liquor), and minerals. Special nutri must frequently be added to stimulate e

duction. This is particularly true of adaptive enzymes where the substrate must be included. Pectic compounds, for example, must be included if certain pectinases are desired.

Maximum enzyme level is obtained after a fermentation period ranging from 18 hours to 7 days, depending on the system. Every possible precaution is exercised during this period to exclude contaminants and maintain a pure culture.

Most industrially produced enzymes are secreted into the medium (hydrolytic enzymes). A few, like glucose oxidase, catalase, and invertase, remain within the cells and must be recovered by autolysis or mechanical fragmentation of the cells.

Bacterial amylase and protease production by *Bacillus subtilis* is a typical example of a submerged fermentation. A medium consisting of starch or sugar, corn steep liquor, proteinaceous material, and inorganic salts is employed. The fermentation proceeds under active aeration 36–48 hours at 34–37°C. The enzymes may be recovered directly from the medium. Relative proportions of amylase and protease may be varied by changes in the mash and growth conditions. A similar process is employed by some for the production of fungal amylase and protease from *Aspergillus oryzae*.

Table 1. Microbial sources of some enzymes

Source	Enzymes	Microorganism
Bacterial	Amylases, proteases	*Bacillus subtilis*, *B. mesentericus*
Fungal	Amylases, amyloglucosidase	*Aspergillus oryzae*, *A. niger*, *A. flavus*
	Proteases	*A. oryzae*, *A. flavus*
	Pectinases	*A. niger*
	Glucose oxidase, catalase	*Penicillium notatum*, *A. niger*
Yeast	Invertase	*Saccharomyces cerevisiae*

Production in semisolid culture. This process has traditionally been employed for the production of amylases and proteases by *A. oryzae*. The current trend is toward greater use of the submerged process.

Wheat bran is the main constituent of the medium. Additional carbohydrates, proteins, minerals, and buffering substances may be added, depending on the enzyme system being produced. The mixture is moistened with water, sterilized with steam, and inoculated with a pure culture of a selected strain of the microorganism.

The moist bran is incubated either in thin layers on trays, or in a slowly revolving horizontal drum. Cool, moist air must be circulated over the bran to maintain proper temperature. Temperatures close to 30°C are usually employed for *A. oryzae* amylase and protease production. Occasional spraying of water over the material may be required to help maintain proper moisture level. The pH may be controlled by inclusion of buffers in the medium or addition of acid or alkali to the spray water. A suitable enzyme level is achieved in 24–48 hours with *A. oryzae*. Other systems may require up to 7 days' incubation. The enzyme systems may be extracted from the "moldy bran" with water. Subsequent recovery procedures are similar to those employed for the submerged fermentation process.

Enzyme recovery and standardization. Cell debris and insoluble mash ingredients are removed by centrifugation or filtration. Diatomaceous earth is usually employed to aid filtration. Buffers and stabilizers are added and the filtrate concentrated in low-temperature vacuum evaporators. Such concentrates may be standardized and marketed di-

Table 2. Typical enzyme applications

Enzyme type and source	Application	Industry
Amylase		
Malt, fungal	Flour supplementation	Baking
Fungal	Bread baking	Baking
Malt	Mashing	Brewing, distilling
Fungal, malt	Precooked baby food	Food
Malt, fungal	Mashing	Industrial alcohol
Bacterial	Starch coatings	Paper
Fungal	Syrup manufacture	Food
Bacterial	Cold-swelling laundry starch	Starch
Fungal, malt, pancreas	Digestive aid	Pharmaceutical
Bacterial, malt, pancreas	Desizing	Textile
Protease		
Fungal	Bread baking	Baking
Pancreas	Prevention of oxidized flavor	Dairy
Pancreas	Protein hydrolysates	Food
Bacterial	Spot removal	Dry cleaning
Pancreas, bacterial	Bating	Leather
Papain, bacterial	Meat tenderizing	Meat
Pancreas, pepsin, papain	Digestive aids	Pharmaceutical
Papain, pepsin	Stabilization	Beer
Rennin	Cheese	Dairy
Trypsin, streptokinase, streptodornase	Wound debridement	Pharmaceutical
Bacterial, pancreas	Desizing	Textile
Invertase		
Yeast	Soft-center candies	Candy
Glucose oxidase (+ catalase)		
Fungal	Glucose removal, oxygen removal	Food
Glucose oxidase (+ peroxidase)		
Fungal	Test paper for diabetes	Pharmaceutical
Pectinase		
Fungal	Pressing, clarification	Wine, fruit juice

rectly. Bacterial amylase for textile desizing is an example of such a product.

Where dry enzymes of a higher degree of purity are desired, the enzymes are recovered by precipitation from the filtrate. Solvents, such as methanol and isopropanol, as well as salts, such as ammonium sulfate, are commonly employed. Techniques of adsorption, dialysis, fractional precipitation, and crystallization may be employed where pure enzymes are required. The requirements of the application determine the degree of purification employed. Most enzymes are marketed as naturally occurring mixtures. Minor constituents of the complex frequently play a significant role in the applications. Inert materials, such as salts, starch, and flour, are employed to adjust the product to the level desired for application. Each lot of enzyme must be assayed to ensure uniform activity of the final product.

Application. Table 2 lists a few typical enzyme applications. Numerous applications are not included because of space limitations. New uses and new industrial enzymes are being developed.

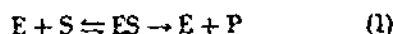
[E.J.B.]

Bibliography: L. A. Underkofler and R. J. Hickey (eds.), *Industrial Fermentations*, vol. 2, 1954.

Enzyme inhibition

The prevention of an enzymic process as a result of the interaction of some substance with an enzyme, so as to decrease the rate of the enzymic reaction. The substance causing such an effect is termed an inhibitor. Enzyme inhibitors are important as chemotherapeutic agents (*see* CHEMOTHERAPY), as regulators in normal control of enzymic processes in living organisms, and as useful agents in the study of biochemistry.

Classification of inhibitors. Inhibitors have been classified as competitive, noncompetitive, or uncompetitive. The following equations show the interactions which may occur, where E represents the free enzyme; S, the substrate; P, the product; I, the inhibitor; ES, the enzyme-subtrate complex; EI, the enzyme-inhibitor complex; and ESI, the enzyme-inhibitor-substrate complex



$$E + S \rightleftharpoons ES \rightarrow E + P \qquad (1)$$



$$E + I \rightleftharpoons EI \qquad (2)$$

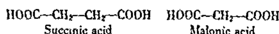


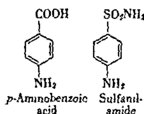
$$ES + I \rightleftharpoons ESI \qquad (3)$$

The interaction of enzyme and substrate is shown in the equation of sequence (1). An inhibitor may interact as indicated in Eqs. (2) and (3) to decrease the rate of sequence (1). A competitive inhibition results when an inhibitor interacts reversibly only with the free enzyme and not with the enzyme substrate complex, as in Eq. (2) but not Eq. (3). A noncompetitive inhibition results from the interaction of an inhibitor, not only with the free enzyme, but also with the enzyme-substrate complex, as in Eqs. (2) and (3). An uncompetitive inhibition is the result of the interaction of an inhibitor with the enzyme-substrate complex, but not with the free enzyme, shown in Eq. (3) but not Eq. (2). These types of inhibitors can be distinguished from each other by a study of their effects on the rate of enzymic processes at different substrate concentrations. The effect of a competitive inhibitor, which binds only free enzyme, can be reversed by sufficiently increased substrate concentrations, so that essentially all of the enzyme is bound into an enzyme-substrate complex. Since both noncompetitive and uncompetitive inhibitors interact with the enzyme-substrate complex, their effects are not nullified by increased concentrations of substrate. An uncompetitive inhibitor exerts relatively less effect at dilute than at more-concentrated substrate concentrations, since less of the enzyme is in the form of the enzyme-substrate complex, with which it interacts. A noncompetitive inhibitor, which reacts with both free enzyme and the enzyme-substrate complex, exerts comparable effects at all substrate concentrations. Special types of inhibitions are also known to occur, such as an irreversible combination of an inhibitor with the free enzyme, but not with the enzyme-substrate complex. The latter type of inhibition results in the substrate's delaying competitively the inactivation of the enzyme.

An early use of noncompetitive inhibitors was in blocking one of a sequence of enzymic reactions, so as to allow accumulation of an identifiable intermediate, thereby elucidating the steps involved in the over-all process. Many of the noncompetitive inhibitors possess active radicals, which interact with certain chemical groups present in enzymes. A similar type of interaction has been shown to occur in certain uncompetitive inhibitions, for example, an interaction of the inhibitor with the reduced, but not the oxidized, form of an enzyme involved in hydrogen transport.

Antimetabolites. Frequently, compounds structurally related to the substrate may form a complex with the enzyme, analogous to the normal enzyme-substrate complex. Products of enzymic reactions have long been known to inhibit the reaction to a greater extent than would be expected by mass action alone, and frequently analogs of the substrate competitively inhibit an enzyme. Malonic acid was found to inhibit competitively the dehydrogenation of succinic acid by succinic acid oxidase, and in 1940, p-aminobenzoic acid was





found to reverse competitively bacterial growth inhibition by the sulfonamide drugs (*see* SULFA DRUGS).

These results led to the recognition that compounds structurally related to essential metabolites can act as competitive inhibitors of the utilization of cell metabolites.

Antimetabolite is a term which has been given to analogs of metabolites that inhibit the utilization of their corresponding metabolite. While most antimetabolites are competitive inhibitors, some appear to react with the enzyme in a manner which results in a noncompetitive inhibition. Antimetabolites are frequently termed metabolite antagonists. Structural analogs of most of the known metabolites have been prepared and found to exert inhibitory effects on the utilization of their corresponding metabolites in many biochemical systems; for example, analog antagonists of amino acids, purines, pyrimidines, vitamins, hormones, and other metabolites have been prepared. Also, metal ion antagonisms have been repeatedly observed. Features necessary for a metabolite analog to be an efficient antagonist of the metabolite include certain groups required in forming complexes with the enzyme, certain specific distances between the necessary groups, certain structural and configurational requirements of the molecule, and other factors, such as similarity in degree of hydration.

Chemotherapy application. Metabolite antagonists have shown some promise, in a few instances, as chemotherapeutic agents. Essential features for chemotherapeutic agents include specificity of the antimetabolite for inhibition of the infective agent but not the host, as well as other features, such as ability to maintain blood levels. Specificity of the sulfonamides probably results from the fact that in the host the same coenzyme is synthesized from folic acid, but not from *p*-aminobenzoic acid, while in the infective organism it is synthesized from *p*-aminobenzoic acid, but not from folic acid Finally, no common intermediate in coenzyme formation is exchanged between the host and infecting organism.

Another application of chemotherapy has been the attempt to produce antitumor agents which will inhibit growing cells, without damage to nonproliferating cells. A number of antimetabolites have shown antitumor effects, but also produce toxicities in the host.

Metabolic pathway demonstration. Inhibitions, involving competition of an analog with a metabolite for an enzyme site, can be used to demonstrate metabolic pathways in living organisms through analysis of the effect of reversing agents other than the metabolite. A secondary reversing agent can theoretically affect a competitive system by (1) altering the concentrations of the competing analog and substrate, or metabolite, at the site of action, or competing per se at the enzyme site; (2) replacing the product of the enzyme, so that inhibition of the particular enzyme is of no consequence to the biological system; (3) sparing the amount of product needed by the system; or (4) increasing the relative number of effective enzyme sites. Methods have been devised to distinguish each general type of reversing agent.

Antagonisms among naturally occurring metabolites can frequently be demonstrated, particularly among the amino acids. Most of these antagonisms involve competitive inhibitions, in which a natural metabolite competes with another metabolite for an essential enzyme, thus inhibiting the enzyme's function. Glutamic acid and aspartic acid are mutually antagonistic in a number of organisms; leucine, isoleucine, and valine are another group of structurally related compounds known to exert mutual antagonisms. Some evidence is also accumulating that antagonisms among naturally occurring compounds may be involved in biological control mechanisms and in the inhibition by a metabolite of its own biosynthesis.

Antibiotics. Antibiotics are in general thought to exert their effect by inhibiting enzyme systems essential for the normal development of the infective agent (*see* ANTIBIOTIC). While in most cases the exact mode of action of antibiotics is not known, some have been shown to have antimetabolite effects. For example, azaserine has been shown to inactivate an enzyme which is involved in the interaction of glutamine with formylglycineamide ribotide, a step in purine biosynthesis. Glutamine competitively delays the azaserine inactivation, indicating not only a competition for an enzyme site but also a further interaction of the enzyme-inhibitor complex to denature the enzyme. *See* BIOCHEMISTRY; ENZYME. [W.S.]

Eocanthocephala

An order of the Acanthocephala characterized by the presence of a small number of giant subcuticular nuclei which are similar to the embryonic nuclei. Body spines may or may not be present and the chief lacunar vessels are dorsal and lateral. The cement gland of the male is a single syncytial organ with a specialized cement reservoir. Ligament sacs persist in the female. Proboscis hooks are few in number and arranged in circles. Eggs are ellipsoidal and thin shelled. These worms are parasitic in cold-blooded vertebrates (turtles, fish). The cystacanth occurs in crustaceans.

Neoechinorhynchus emydis is a parasitic species in North American turtles. The body is white and typically curved in extended specimens. The females are 10–22 mm long whereas the males are 6–14 mm. The body is 0.5–1 mm wide. There are typically 6 giant subcuticular nuclei of which 5 are middorsal and 1 midventral. The male reproductive organs are in the posterior half of the body, the testes are elongate, contiguous in tandem, and the large syncytial cement gland contains 8 nuclei. The proboscis is globular with 3 circles of 6 hooks, each in quincuncial arrangement. The lemnisci are unequal in length; the longer contains 2 nuclei, the shorter only 1 nucleus. Eggs are oval with 3 membranes; the middle membrane does not completely

encircle the acanthor. The first intermediate host is an ostracod (*Cypria maculata*); the second intermediate host is a snail (*Campeloma rufrum*), with the cystacanth encysting in the foot of the snail. The definitive host is the map turtle *Graptemys geographica*. In this life cycle the snail appears to be an essential second intermediate host and not simply a transport host. *See* ACANTHOCEPHALA. [D.V.M.]

Eocene

The next to the oldest of the five major world-wide divisions (Epochs) of the Tertiary Period (Cenozoic Era); the epoch of geologic time extending from the end of the Paleocene to the beginning of the Oligocene; the middle epoch of the older Tertiary (Paleogene or Nummulitic). *See* CENOZOIC; TERTIARY.

PRE-CAMBRIAN		PALEOZOIC							MESOZOIC			CENOZOIC	
						CARBONIFEROUS							
ARCHEOZOIC (EARLY PRECAMBRIAN)	PROTEROZOIC (LATE PRECAMBRIAN)	CAMBRIAN	ORDOVICIAN	SILURIAN	DEVONIAN	Mississippian	Pennsylvanian	PERMIAN	TRIASSIC	JURASSIC	CRETACEOUS	TERTIARY	QUATERNARY

Tertiary					Quaternary	
Paleocene	Eocene	Oligocene	Miocene	Pliocene	Pleistocene	Recent

Sir Charles Lyell, the British geologist, in 1833 subdivided the Tertiary into Pliocene (youngest), Miocene, and Eocene. This usage of the term Eocene included those segments of the Tertiary which are now classified as Paleocene (oldest), Eocene, and Oligocene; as such, it is more or less synonymous with older Tertiary (Paleogene or Nummulitic).

Subsequent to the proposal of Lyell's original subdivisions, it was realized that certain contemporaneous deposits were being assigned by some workers to the upper Eocene and by others to the lower Miocene. As a result of studies of these particular strata in western Europe, E. Beyrich in 1854 proposed the name Oligocene and indicated that it was intermediate between the older Eocene and the younger Miocene. Twenty years later, in 1874, as a result of studies of Eocene fossil plants, W. Schimper introduced the term Paleocene for strata considered at that time to be the basal part of the Eocene but which contain plants that are sufficiently different from those higher in the Eocene to justify their separation. The ensuing three-quarters of a century have witnessed the general acceptance of Paleocene, a restricted Eocene, and Oligocene as world-wide divisions of the older Tertiary. *See* OLIGOCENE; PALEOCENE.

The Eocene Series includes all rocks formed during the Eocene Epoch, but the term is used most specifically with respect to the sedimentary rocks formed during this portion of the Tertiary Period. These latter contain the plant and animal remains which are the primary bases for the identification of Eocene age.

Although the limits of the Eocene are marked in places by physical breaks and by cyclic sequences of transgressive (invading) and regressive (withdrawing) marine to marginal marine deposits in the rock record, such physical phenomena are not present everywhere; thus differentiation of Eocene from the overlying and underlying divisions is often difficult. In such instances identification of Paleocene, Eocene, and Oligocene is primarily a comparative paleontologic problem.

The strata. Eocene strata include all the common sedimentary types, varying from marine through marginal-marine or intermediate to terrestrial in origin. They are typically unconsolidated to poorly consolidated and are widely dispersed throughout the world. The terrestrial beds are best known in the continental interiors, while the marginal and marine beds are most widespread near the continental margins in the areas of the coastal plains and continental shelves. Especially noteworthy examples of these various strata are known in (1) the Gulf Coastal province of the United States and Mexico; (2) the Pacific border of North America; (3) the intermontane basins of the North American Cordillera; (4) parts of England, France, and Belgium; (5) the Mediterranean region of southern Europe and northern Africa; (6) the area around the Persian Gulf; and (7) the Himalaya Mountains.

The classical areas for study of Eocene strata are the regions around London and Paris in Europe and the Gulf Coast of the United States. Many eminent paleontologists have studied the rich marine Eocene faunas in these areas. They have been largely responsible for subdividing the Eocene Series into numerous units (stages) which in only a few instances can be correlated with certainty from one geologic province to another. A second major paleontologic problem is correlating terrestrial, mammal-bearing Eocene sediments (as in the Alps and the Rocky Mountains) with marine, invertebrate-bearing Eocene deposits (as in the Paris Basin and the Gulf Coast).

Most of the known marine and marginal Eocene strata are relatively undisturbed and flat lying and occur near sea level, but crustal [illegible] are appreciably deformed some and some have [illegible] uplifted to considerable elevations. Others [illegible] been depressed or have subsided below sea [illegible]. The terrestrial strata were deposited [illegible] level in the continental areas and for the [illegible] have remained in this general position. [illegible] thicknesses of Eocene deposits vary from [illegible] to considerably more than 10,000 ft. [illegible]

important quantities of oil and gas, potable ground water, clay, sand, marl, lignite, phosphate, and low-grade iron ores in various parts of the world.

Igneous rocks of Eocene age include both intrusive and extrusive types, the latter being best known because they commonly occur as layered volcanic materials at the surface and are therefore more readily available for observation. Igneous activity, generally, was not as widespread in the earlier Tertiary (Paleogene or Nummulitic) as in the later Tertiary (Neogene); consequently Oligocene, Eocene, and Paleocene igneous rocks are not as abundant as those of Miocene and Pliocene age.

Fauna. During the Eocene, mammals initiated all of the main evolutionary lines which have culminated in the present terrestrial fauna. Hoofed mammals are represented in Eocene deposits by small, multitoed horses (*Eohippus*) and larger rhinoceroslike creatures (titanotheres); carnivores are represented by creodonts; and rodents and primates are also represented by primitive forms. Whales made their first appearance in the Eocene seas. The Indo-Pacific tropical marine pelecypods, gastropods, echinoids, and foraminifers, which originated in the Paleocene, spread progressively westward across the Mediterranean region and reached the Americas in late Eocene time. *See* PALEOBOTANY; PALEONTOLOGY. [A.H.CH.; C.E.M.]

Eocrinoidea

A class of extinct Pelmatozoa that had biserial brachioles like those of cystoids, combined with a theca like that of crinoids. *Cryptocrinus*, a representative genus, had an ovoid theca composed of three rings of plates and a tegmen which carried five pairs of brachioles. Their affinities have been sought among both cystoids and crinoids. Eocrinoidea appeared in the Lower Cambrian and are not known later than the Middle Ordovician. *See* CRINOIDEA; CYSTIDEA; PELMATOZOA. [H.B.F.]

Ephaptic transmission

Electrical transfer of activity to a postephaptic unit by the action current of a preephaptic cell. This phenomenon is analogous to local circuit excitation within a single electrically excitable cell. Membrane of relatively low resistance at ephaptic junctions to permit current flow sufficient for excitation is the requisite. Different segmental portions of septate giant axons of worms and crustaceans form unpolarized ephaptic junctions, and current flow from either side of the junction excites the other cell. Polarized ephaptic transmission, resembling that at synapses, occurs at the junction between cord and root giant axons of crayfish. An impulse in the former sends excitatory current into the latter, but resistance is high to current flow in the opposite sense. This is an extreme example of rectifier action, found to a less pronounced degree in other membranes. Field currents, resulting from the activity of masses of cells in the central nervous system, may act on electrically excitable nerve fibers in place of or as a supplement to postsynaptic potentials. Ephaptic junctions probably also occur between the neurons of cardiac ganglion in lobsters and some fishes. *See* SYNAPTIC TRANSMISSION. [H.GT.]

Ephedra

A genus of low, leafless, green-stemmed shrubs belonging to the plant order Gnetales of the class Gymnospermae. They grow in dry, alkaline soils around the world. In the southwestern part of the United States these plants are called Mormon tea and jointfir. The drug ephedrine is extracted from the Asiatic species, *Ephedra sinica* and *E. equisetina*. It is used medicinally in the treatment of colds, hay fever, and asthma. *See* EMBRYOPHYTA; GNETALES. [P.D.S.]

Ephemeris

A table of data, especially astronomical data, that depend on the time, usually arranged with values of the time in the left-hand column. A lunar ephemeris, for example, may give the right ascension and declination of the Moon for every hour of a particular year. *The American Ephemeris and Nautical Almanac* is an annual volume published by the Nautical Almanac Office, U.S. Naval Observatory, containing ephemerides of the Sun, Moon, planets, and satellites, with many other astronomical data. *See* NAUTICAL ALMANAC. [G.M.C.]

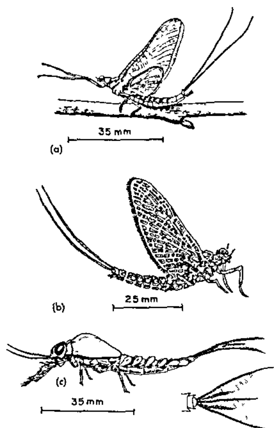
Ephemeroptera

An order of insects commonly known as mayflies. They are aquatic insects that live in clean, fresh waters during their immature, or nymphal, lives. Nymphs are adapted to aquatic situations ranging from ponds to mountain streams. Mayflies are complete vegetarians and, in one or another stage of life, are present throughout the year. They constitute a year-round basic food supply for carnivores of their communities, especially for fishes, trout in particular. Nymphs and adults are models for artificial bait flies.

Mayflies drop their eggs in packets into the water, or scatter them over its surface. Certain small species fix the eggs to surfaces of rocks partly submerged in rapid currents. The nymphal stage lasts from a few weeks to over two years, according to the species. At the end of their growth, the nymphs rise to the surface, shed their skins, and become subadults, or subimagoes. Within a short time, which varies from minutes to a day or two, they shed the subadult skin and become adults, or imagoes.

Mayfly nymphs are distinguished from all other aquatic insects by the paired tracheal gills present on the back of each of the first seven abdominal segments. The body ends in two or three finely segmented tailpieces. The mouthparts are highly interesting examples of adaptations for scraping, cutting, and crushing the various plant cells, which are the only food of these insects.

As subadults, mayflies have the general adult form and useless mouthparts. They are clothed by



Ephemeroptera. (a) *Mayfly nymph, Chirotenetes sp., from rapid stream (A. H. Morgan, Field Book of Ponds and Streams, Putnam, 1930), (b) Mayfly, subadult male, Callibaetis sp., newly emerged from water, standing on surface film (A. H. Morgan, Kinships of Animals and Man, McGraw-Hill, 1955); (c) Mayfly, adult male, Hexagenia sp., in usual resting position (A. H Morgan, Field Book of Ponds and Streams, Putnam, 1930).*

a thin, furry, grayish skin which accounts for the name duns. The location of the subadult stage, however, between the nymphal and adult stages, suggests the sequence of stages in the metamorphosis of other insects such as larva, pupa, and adult of the butterfly. The subimago stage ends with the final molt of the mayfly's life. No other insects molt in their winged form.

The adult emerges with transparent wings, shining body, and useless mouthparts; the males have extremely large eyes and long front legs The quivering of the tailpieces during the mating flight gave them the name spinners. Thousands join in the twilight mating swarms, rhythmic dancing flights unequalled by other insects *See* INSECTA [A H.M.]

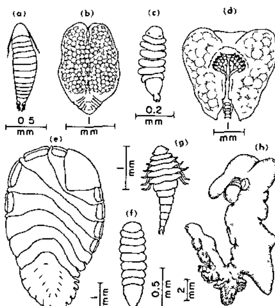
Bibliography: A. H. Morgan, *Field Book of Ponds and Streams*, 1930; J. G. Needham, J. R. Traver, Y. C. Hsu, *Biology of Mayflies*, 1935.

Epicaridea

A suborder of the Isopoda which are parasitic on various crustaceans, mainly marine forms. The females are sometimes modified so strongly as to leave almost no indication of their isopod nature. The female pierces the host's skin with the aid of styliform mandibles to suck blood. It seizes the host's skin by means of its prehensile pereiopoda. The pleopoda are respiratory in function. The dwarf male attaches to the female, from which it takes nourishment, and retains its isopod structure. The first larva, the epicaridium, is free-living, while the second, the microniscium, is temporarily ectoparasitic on Copepoda. The third larval stage, the cryptoniscium, is free-swimming and seeks a final host, on which it undergoes final development to the adult. The suborder is divided into two tribes, the Cryptoniscini and Bopyrini.

Cryptoniscini. These crustaceans live on Entomostraca and are protandrous hermaphrodites. The cryptoniscium larva attaches to a suitable female, develops testes without any change in its external appearance, and acts as a male After the death of the host, it metamorphoses into a female, ultimately undergoing anatomical degradation that leaves it a mere bulky sac distended with developing embryos. The brood chamber is internal and arises from paired invaginations of the sternum *Hemioniscus* lives in the mantle cavity of Cirripedia, *Cyproniscus* lives in the shell of Ostracoda, and *Liriopsis* infests Rhizocephala which is, in turn, a parasite of crabs *Asconiscus*, *Podascon*, and *Cabirops* live in the brood pouch of Mysidacea, Amphipoda, and Isopoda respectively.

Bopyrini. These isopods are dioecious, but it has been shown experimentally, in some forms, that the cryptoniscium is ambipotent, and a presumptive young male removed from the female develops into either a female or intersex. The tribe includes



Epicaridea. (a) *Cyproniscus ovalis* Shiino, male; (b) female (c) *Prodajus bilobatus* Shiino, male; (d) female. (e) *Bopyrus squillarum* Latreille, female; (f) male. (g) *Portunion flavidus* Shiino, male; (h) female.

important quantities of oil and gas, potable ground water, clay, sand, marl, lignite, phosphate, and low-grade iron ores in various parts of the world.

Igneous rocks of Eocene age include both intrusive and extrusive types, the latter being best known because they commonly occur as layered volcanic materials at the surface and are therefore more readily available for observation. Igneous activity, generally, was not as widespread in the earlier Tertiary (Paleogene or Nummulitic) as in the later Tertiary (Neogene); consequently Oligocene, Eocene, and Paleocene igneous rocks are not as abundant as those of Miocene and Pliocene age.

Fauna. During the Eocene, mammals initiated all of the main evolutionary lines which have culminated in the present terrestrial fauna. Hoofed mammals are represented in Eocene deposits by small, multitoed horses (*Eohippus*) and larger rhinoceroslike creatures (titanotheres); carnivores are represented by creodonts; and rodents and primates are also represented by primitive forms. Whales made their first appearance in the Eocene seas. The Indo-Pacific tropical marine pelecypods, gastropods, echinoids, and foraminifers, which originated in the Paleocene, spread progressively westward across the Mediterranean region and reached the Americas in late Eocene time. *See* PALEOBOTANY; PALEONTOLOGY [A.H.CH; C.E.M.]

Eocrinoidea

A class of extinct Pelmatozoa that had biserial brachioles like those of cystoids, combined with a theca like that of crinoids. *Cryptocrinus*, a representative genus, had an ovoid theca composed of three rings of plates and a tegmen which carried five pairs of brachioles. Their affinities have been sought among both cystoids and crinoids. Eocrinoidea appeared in the Lower Cambrian and are not known later than the Middle Ordovician. *See* CRINOIDEA; CYSTIDEA, PELMATOZOA. [H.B.F.]

Ephaptic transmission

Electrical transfer of activity to a postephaptic unit by the action current of a preephaptic cell. This phenomenon is analogous to local circuit excitation within a single electrically excitable cell. Membrane of relatively low resistance at ephaptic junctions to permit current flow sufficient for excitation is the requisite. Different segmental portions of septate giant axons of worms and crustaceans form unpolarized ephaptic junctions, and current flow from either side of the junction excites the other cell. Polarized ephaptic transmission, resembling that at synapses, occurs at the junction between cord and root giant axons of crayfish. An impulse in the former sends excitatory current into the latter, but resistance is high to current flow in the opposite sense. This is an extreme example of rectifier action, found to a less pronounced degree in other membranes. Field currents, resulting from the activity of masses of cells in the central nervous system, may act on electrically excitable nerve fibers in place of or as a supplement to postsynaptic potentials. Ephaptic junctions probably also occur between the neurons of cardiac ganglion in lobsters and some fishes. *See* SYNAPTIC TRANSMISSION. [H.GT.]

Ephedra

A genus of low, leafless, green-stemmed shrubs belonging to the plant order Gnetales of the class Gymnospermae. They grow in dry, alkaline soils around the world. In the southwestern part of the United States these plants are called Mormon tea and jointfir. The drug ephedrine is extracted from the Asiatic species, *Ephedra sinica* and *E. equisetina*. It is used medicinally in the treatment of colds, hay fever, and asthma. *See* EMBRYOPHYTA; GNETALES. [P.D.S.]

Ephemeris

A table of data, especially astronomical data, that depend on the time, usually arranged with values of the time in the left-hand column. A lunar ephemeris, for example, may give the right ascension and declination of the Moon for every hour of a particular year. *The American Ephemeris and Nautical Almanac* is an annual volume published by the Nautical Almanac Office, U.S. Naval Observatory, containing ephemerides of the Sun, Moon, planets, and satellites, with many other astronomical data. *See* NAUTICAL ALMANAC. [G.M.C.]

Ephemeroptera

An order of insects commonly known as mayflies. They are aquatic insects that live in clean, fresh waters during their immature, or nymphal, lives. Nymphs are adapted to aquatic situations ranging from ponds to mountain streams. Mayflies are complete vegetarians and, in one or another stage of life, are present throughout the year. They constitute a year-round basic food supply for carnivores of their communities, especially for fishes, trout in particular. Nymphs and adults are models for artificial bait flies.

Mayflies drop their eggs in packets into the water, or scatter them over its surface. Certain small species fix the eggs to surfaces of rocks partly submerged in rapid currents. The nymphal stage lasts from a few weeks to over two years, according to the species. At the end of their growth, the nymphs rise to the surface, shed their skins, and become subadults, or subimagoes. Within a short time, which varies from minutes to a day or two, they shed the subadult skin and become adults, or imagoes.

Mayfly nymphs are distinguished from all other aquatic insects by the paired tracheal gills present on the back of each of the first seven abdominal segments. The body ends in two or three finely segmented tailpieces. The mouthparts are highly interesting examples of adaptations for scraping, cutting, and crushing the various plant cells, which are the only food of these insects.

As subadults, mayflies have the general adult form and useless mouthparts. They are clothed by

Epidemiology

The study of the mass aspects of disease. The word epidemic literally means "upon the people," and was coined to describe the way an infectious disease, in spreading through a group of people, gives the impression of an affliction that has been placed upon the community as a whole. An observer reflecting on such an epidemic would ask questions about the circumstances which allowed the disease to develop and flourish, what permitted some members to escape, and what brought the epidemic to an end. These are questions about the disease as it affected groups of people, rather than individuals in the group, and they are the typical problems of epidemiology. *See* EPIDEMIC.

In recent years diseases other than infectious diseases have been included in the subject matter of epidemiology. In the United States in 1958 approximately 100,000 people died of accidents, and about 250,000 from cancer. It is quite natural therefore, that these massive noninfectious diseases should be studied as group phenomena, and one should speak of the epidemiology of accidents, cancer, heart disease, or mental disease. For both infectious and noninfectious diseases there are similar general problems such as identifying cases in the population at risk, and finding which circumstances lead to an increase and which to a decrease in the disease. In spite of the differences in detail arising in a study of Asian influenza and a study of suicide in a particular population, both are fundamentally investigations of disease rates and a comparison of the way these rates vary from one subgroup to another; for example, is there a difference in the rates for the two sexes, the various age groups, or different socioeconomic classes? A consideration is made of a number of rates that have proved useful in the study of epidemiology. Some of the most important are the mortality rate, the morbidity rate, the prevalence ratio, the incidence rate, and the case fatality ratio.

Mortality rate. The mortality rate for a given year is the number of deaths occurring in the year per 1000 total midyear population. In the United States in 1946 there were 1,395,600 deaths and the population as of July 1 was 139,893,000. Hence the death rate was

$$\frac{1{,}395{,}600 \times 1000}{139{,}893{,}000} = 10.0$$

This rate is sometimes referred to as the crude death rate because no account is taken of such factors as the age and sex of the members of the population. Since France has a higher percentage of old people than the United States, and Israel a lower percentage, it is not surprising to find that the crude death rate in France is considerably higher than that of the United States and the death rate in Israel considerably lower. These objections to the crude death rate can be met by computing separate rates for particular age groups in each sex, for example, the rate for males aged 15–24, another rate for males 25–34, and so on. The process can be carried further by computing separate death rates for individual diseases such as tuberculosis, or groups of diseases such as diseases associated with child bearing. The rates for particular age groups, or particular causes of death, or specified sex are known as specific rates. They are often essential for the thorough study of an epidemic. However, it is true in general that the more specific a rate, the more difficult it is to collect accurately the information necessary for the computation. Thus to compute the death rate from accidents in 1959 among male children aged 5–14 years, it is necessary to know both the number of deaths from accidents among the group in 1959 and also what is called the population at risk, that is, the number of boys in the given age group at the time and place under consideration. The difficulties of getting this information are such that the more reliable, though less specific, crude rate still has a useful place in the methodology of the subject. An example of the effect of a major epidemic on this rate is shown by the death rate for the United States during the influenza epidemic of 1918 The figure for that year was 18.1 deaths per 1000 population; for the previous year it was 14.0, and for the succeeding year 12.9.

Morbidity rates. There are certain diseases which cannot be studied by a means of mortality rates of any kind. Mental diseases, arthritis, and the common cold are examples of conditions the epidemiology of which is reflected inadequately by death rates. A more useful approach is to consider morbidity rates which give the number of cases of disease per 100,000 population, either as cases existing at a particular point in time or as cases occurring in a particular period of time. Knowledge of these rates for the population in general is inadequate. There are many countries where the reporting of the fact of death is virtually 100%; but there is no comparable record by which one can ascertain how many people suffer from disease. It is true that there is compulsory reporting in many countries of all discovered cases of certain infectious diseases such as smallpox and leprosy; that schools, hospitals, factories, and health insurance groups have records giving useful information about the diseases found in special segments of the population; and that morbidity surveys such as the current United States National Health Survey, based on household inquiry by skilled interviewers, have proved useful. But, in general, whenever the information about morbidity covers the whole population or a representative sample of it, the data are limited in detail. Thus the course of the epidemic of Asian influenza in the United States in 1957–1958 can be traced accurately, as to the number of cases involved, from the returns of the National Health Survey, but there is no information on the finer details about the disease, such as can be studied when samples of blood are availa[illegible] special examinations.

Disease rates. In the investigation of a particular epidemic, ascertainment of disease rates is of great importance. It is necessary to distinguish two main kinds of rates, the one based on the incidence of the disease and the other on its prevalence.

Incidence rate. The incidence rate is concerned with the cases of disease that develop or, more strictly, that come under diagnosis in a given period of time. The incidence rate per 100,000 population for a given year is therefore

$$\text{Incidence rate} = \frac{\text{number of new cases occurring during year}}{\text{midyear population}} \times 100{,}000$$

This rate measures the risk of developing the disease in the given time period; for convenience it is expressed in terms of so many cases per 100,000 population at risk.

Prevalence. The prevalence ratio is concerned with the cases of disease that exist at a particular point in time. The concept is clearest when the point of time is thought of as being instantaneous; for example, one might ask how many cases of tuberculosis per 100,000 population existed at a specified moment, and this would then be a prevalence ratio for the disease. Prevalence ratios are also computed, especially for chronic diseases, ver periods as long as a year. The essential char- eristic is still maintained, however, that the unt is based on cases existing at some part of the given period, and not necessarily developing within the given period. The prevalence ratio takes in cases that developed prior to the period under consideration, provided the disease persisted until the prevalence count was made.

In the case of diseases which last for several years, such as tuberculosis and diabetes, the prevalence may be much greater than the incidence. In fact, there is some evidence that the modern effective treatment of diabetes, which holds the disease in check without curing it in a fundamental sense, *has actually increased the* prevalence of the disease. A similar point arose during World War II in connection with the epidemiology of tuberculosis. Since it is well known that in Tennessee the death rate from tuberculosis is higher among the Negroes than the whites, it was expected that in the Selective Service Examinations a higher prevalence of the disease would be found in Negroes than in whites. This was not so, presumably because for various reasons the disease ran a longer but less fatal course in the whites.

The distinction between incidence and prevalence is crucial for much epidemiological work. In particular, if one wishes to decide whether a factor is associated with a particular disease, for example, baldness with disease of the coronary arteries, one should make the correlation on cases that arise in a given period of time. The cases of disease that exist at some point in time represent a biased sample of the total, since they include only those who have survived for a period of time, and if survival were better for bald men than for the rest, one would be in danger of concluding that baldness predisposed to coronary artery disease even if, in fact, this were not true.

During the period an incidence rate is being measured, a person may have more than one attack or may suffer from more than one disease. Consequently it is necessary to state whether the unit of measurement is a person, an episode of illness, or a particular disease; the count will vary according to the definition. There is the further obstinate problem of defining when a person is sick. For some purposes this may be done by the subject himself, in terms of the number of days of absence from work. In other instances, especially when long-lasting diseases such as diabetes are involved, periodic examinations may be made by a physician. The problem of the careful definition of cases is important in epidemiological research and this has led to a close cooperation in modern work between physicians, who identify cases by clinical means; laboratory workers, who identify cases by microbiological, biochemical, or other laboratory examinations; and statisticians, who make the comparisons of the disease rates in various subgroups of the population at risk.

Case fatality ratio. In addition to studying the presence or absence of a disease, the epidemiologist often uses some measure of the severity of the disease. One index of severity is the case-fatality ratio which is defined as the number of deaths expressed as a percentage of the number of cases.

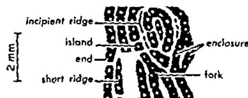
Experimental epidemiology. The observational methods so far discussed do not exhaust the possibilities of studying the factors influencing epidemics. The experimental approach has also been used and experimental epidemiology has great potential for improving knowledge of the subject. An epidemic is started by introducing into an animal colony a certain number of infected animals. The course of the subsequent epidemic can then be followed and a study made of the influence of factors such as diet, the percentage of susceptible animals, and the intimacy of the contact between the animals. It is easy to show by such methods, for example, that the introduction of fresh susceptibles into the population from time to time has a marked effect in sustaining an epidemic. *See* BACTERIOLOGY, MEDICAL. [C.WH.]

Epidermal ridges

Minute corrugations of the skin. They compose a sculpturing, termed dermatoglyphics, which characterizes the palmar and plantar surfaces of man and the nonhuman primates. These areas lack hair and sebaceous (oil) glands, but sweat glands are numerous. In certain kinds of monkeys, a portion of the under surface of the tail bears similarly specialized skin.

Epidermal ridges do not course in a uniform direction, like the ribs of corduroy, but are arranged in configurations of variable types as exemplified in the familiar arches, loops, and whorls of the

fingers. Ridges are dissimilar to the ribs of corduroy also in that they present frequent interruptions, branchings, and other irregularities. All these details are unchanged in lifetime except as they may be destroyed or damaged by accident or disease. Even a small area of skin has a unique individuality of these detailed features, which provide a basis for positive personal identification.



Characteristics of ridges, as seen in a print. (From H. Cummins and C. P. Midlo, Fingerprints, Palms, and Soles, McGraw-Hill, 1943)

The ridges over the human hand as a whole, in young adult males, average 0.48 mm in breadth. They are slightly narrower in females, 0.43 mm

Ridges serve two functions: (1) They increase security of contact with objects, in the manner of the milling of a tool handle. Ducts of sweat glands open on the summits of ridges, and moistening of the skin augments the security of contact. (2) They enhance the sense of touch In passing the fingers or palm over an object for judging its texture, the slight displacement of ridges heightens stimulation of the underlying nerve endings.

The characteristics of individual ridges and their collective configurations are not ordinarily studied by direct inspection of the skin. Instead, they are examined in prints, usually impressions of the inked surface, made for the purpose of record, or natural imprints made by chance contact. *See* FINGERPRINT; INTEGUMENT; INTEGUMENTARY PATTERNS. [H.CU.]

Epidermis, plant

The outermost layer (occasionally several layers) of cells on the primary plant body. Its origin and structure are variable. In addition to a general description of the epidermis of plants, this article singles out five structural components of the tissue: (1) cuticle; (2) stomatal apparatus (including guard cells and subsidiary cells); (3) bulliform (motor) cells; (4) trichomes; and (5) root hairs

In most roots the epidermis has a common origin with other tissues and therefore becomes established a short distance back of the apex (*see* MERISTEM, APICAL). In shoots that have stratified meristems, the outermost layer (first layer of the tunica) usually gives rise to the epidermis. In many grasses, however, this layer divides and contributes to internal tissues of the leaf primordia before the epidermis is organized In plants that have unstratified shoot meristems, the embryonic epidermis (protoderm) may originate on the flanks of the apical cone or on organ primordia.

Leaves, herbaceous stems, and floral organs usually retain the epidermis through life. Most woody stems retain it for one to many years, after which it is replaced (*see* PERIDERM). In roots it is usually short-lived.

Ordinarily the epidermis is compact and devoid of intercellular spaces, aside from the stomata. In surface view the individual cells commonly appear polygonal or have wavy outlines (Fig. 1). Those on stems and linear leaves may be conspicuously elongated parallel to the axis of the organ. Specialized epidermal cells include the silica and cork cells in the leaves of some grasses, the lithocysts in *Ficus* and *Cannabis* (Fig. 2*a*), and fiberlike cells of *Raphia* and *Stylidium*. In the dry bulb scales of garlic the epidermis consists partly or entirely of heavily thickened sclereids (*see* SCLERENCHYMA). In the seed coats of many legumes it consists of a compact layer of macrosclereids arranged in columnar fashion. The epidermis of certain Polypodiaceae is photosynthetic. There are also plants in which the entire leaf, excluding the veins, consists of a double (*Elodea*) or even a single (*Hymenophyllum*) layer of epidermis.

Differential thickening of the walls characterizes most epidermal cells. In the more delicate types of epidermis, the outer tangential wall is commonly the thickest (Fig. 2*b*). In the leaves of many conifers, however, the walls of epidermal cells are so massively thickened that the cell cavities are almost obliterated (Fig. 2*c*).

Multiple epidermis. This tissue is several layers of cells in depth. In its formation the protoderm undergoes periclinal (parallel with the circumference) divisions A multiple epidermis occurs in many species of *Ficus*, *Begonia*, and *Peperomia*. The outermost layer of a multiple epidermis resembles a uniseriate (single series) epidermis, whereas the other layers constitute a water storage tissue. In some peperomias, the water storage tissue is thicker than the rest of the leaf. The velamen (parchmentlike covering) of orchid roots is a multiple epidermis.

Cutin, cuticle, and waxes. Cutin is a mixture of fatty substances characteristically found in epi-

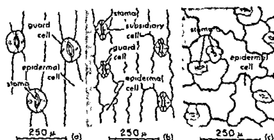


Fig. 1. Epidermis in surface view. (a) Epidermis of leaf of *Iris* showing kidney-shaped guard cells; (b) epidermis of leaf of *Zea* showing dumbbell-shaped cells. There are two subsidiary cells associ[illegible] each stoma; (c) Epidermis from leaf of [illegible] stoma is associated with three or four

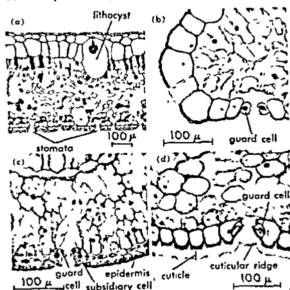


Fig. 2. Specialized epidermal cells in transections of leaves. (a) Leaf of *Ficus elastica* with approximately three layers of cells in the multiple epidermis on each side of the leaf. Lithocyst containing cystolith whose calcium deposits were removed during preservation; (b) Margin of leaf of *Lilium* showing relatively thin-walled epidermis; (c) Leaf of *Pinus* showing thick-walled epidermis and sunken guard cells; (d) Leaf of *Clivia* showing heavy cuticle.

dermal cells. It impregnates the outer cell walls and occurs as a continuous layer (cuticle) on the outer surface (Fig. 2*d*). The cuticle covers the surfaces of young stems, leaves, floral organs, and even apical meristems. The cuticle can be stripped from some organs in relatively large sheets, a feature indicating its continuity. Cutin has been detected in the walls of cells adjacent to the intercellular spaces within leaves. There are some reports that it occurs in the epidermis of young roots.

Waxes appear as a deposit on the outside of the cuticle in many plants; the bloom on purple grapes and plums is an example. Most often the waxes are present in small quantity, but the leaves of some plants may be almost white with wax (*Echeveria subrigida*). The waxes of a few species are of great commercial value in the manufacture of polishes for floors, furniture, automobiles, and shoes. Carnauba wax is obtained from the leaves of a palm, *Copernicia cerifera;* candelilla wax, from the pencil-like stems of *Euphorbia antisyphilitica.* Other substances such as gums, resins, and salts, usually in crystalline form, may be deposited on the outside of the cuticle.

Stomata. The apertures in the epidermis which are surrounded by two specialized cells, the guard cells, are known as stomata. The singular form, stoma, is derived from the Greek word for mouth. However, some authorities prefer to include both aperture and guard cells within the concept of

ium) and on the sporophytes (spore-producing generation) of certain nonvascular plants (mosses, *Anthoceros*). The apertures of stomata are contiguous with the intercellular space system of underlying tissues and thus permit gas exchange between internal cells and the external environment (*see* PLANT PHYSIOLOGY).

The number of stomata per square centimeter of epidermis varies from none to more than 100,000. In many plants stomata are restricted largely to the lower surface of the leaf. Occasionally they occur only on the upper surface (*Castalia, Ammophila*). They may also occur in equal numbers on both sides of the leaf (*Abronia, Atriplex*). In oleander, stomata are restricted to special crypts on the lower leaf surface. In many dicotyledons stomata are arranged at random; in many monocotyledons and in the leaves of gymnosperms they are arranged with their long axes parallel to that of the leaf. The guard cells are frequently associated with subsidiary cells, which differ from other epidermal cells in size and form.

The guard cells of many plants are at about the same level as the other epidermal cells; in others they project above them. Xerophytes, plants which can subsist with a small amount of moisture, often have a thick-walled epidermis and the guard cells may be deeply sunken. In the majority of plants guard cells are kidney-shaped in surface view; in grasses, however, they resemble a pair of dumbbells. The guard cells undergo changes in form during the changes in turgor (water pressure), a phenomenon associated with the characteristic uneven thickening of the guard cell walls.

The opening and closing of the stomatal aperture is caused by relative changes in turgor between the guard cells and surrounding epidermal cells. It is accompanied by hydrolysis, or condensation, of starch within the guard cells (*see* PLANT, WATER RELATIONS OF).

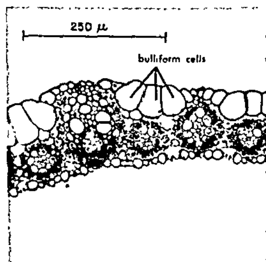


Fig. 3. Transection of leaf of *Andropogon* showing large bulliform cells in the upper epidermis.

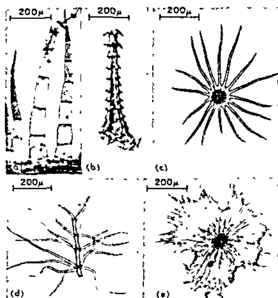


Fig. 4. Trichomes. (a) Uniseriate hairs of *Gynura;* (b) Unicellular anchor hair of *Mentzelia,* (c) Stellate hair of *Solanum;* (d) Branched candelabra hair of *Platanus;* (e) Peltate scale of *Shepherdia.*

Bulliform (motor) cells. These large, highly vacuolated cells occur on the leaves of many monocotyledons but are probably best known in grasses (Fig. 3). They are thought to play a role in the unfolding of developing leaves and in the rolling and unrolling of mature leaves in response to alternating wet and dry periods. Not all investigators, however, subscribe to this idea. The water vesicles in the epidermis of the ice plant, *Mesembryanthemum crystallinum,* are similar in appearance to bulliform cells.

Trichomes. Appendages derived from the protoderm are known as trichomes; the simplest are protrusions from single epidermal cells. Included in the concept, however, are such diverse structures as uniseriate hairs (Fig. 4*a*), multiseriate hairs (*Begonia, Saxifraga*), anchor hairs (Fig. 4*b*), stellate hairs (Fig. 4*c*), branched (candelabra) hairs (Fig. 4*d*), peltate scales (Fig. 4*e*), stinging hairs, and glandular hairs. Many of these structures arise on developing plant organs and are most conspicuous at this stage. They may be persistent or short-lived. Persistent trichomes are responsible for the silver-gray color of sagebrush (*Artemisia*), Spanish moss (*Tillandsia*), and Russian olive (*Elaeagnus*). Mature trichomes may or may not retain their protoplasts. The assignment of functions to trichomes is often uncertain, although they are useful in plant identification. Cotton and kapok fibers are unicellular epidermal hairs. Cotton fibers grow from the seed coat, those of kapok from the pod epidermis.

Root hairs. These are thin-walled extensions of certain root epidermal cells. *See* ROOT (BOTANY). They develop only on growing root tips and may arise from any epidermal cell, or from specialized cells known as trichoblasts Root hairs increase the absorbing area of the root tip manyfold. The life of a given root hair is usually numbered in days. Persistent root hairs, reported for some plants, probably play no role in absorption. *See* PHLOEM; PLANT ORGANS; PLANT TISSUE SYSTEMS; SECRETORY STRUCTURES, PLANT; XYLEM. [N.H.B.]

Bibliography: See PLANT ANATOMY.

Epidiorite

A rock name applied to an altered diorite or gabbro. Gabbroic rocks exposed to mechanical pressure and hydrothermal action will easily alter into green rocks. Many greenstones are in reality epidiorites. The plagioclase decomposes and will be replaced by an aggregate of epidote and albite (saussuritization) and the pyroxene becomes transformed to hornblende (uralitization). Patches of the original rock may still remain and be recognized here and there. The term epidiorite has been generally used to cover rocks of this category. In such rocks the hornblende may retain parting on the crystallographic plane (001) or on (100) inherited from the pyroxene it has replaced, and the saussuritized plagioclase may retain the idiomorphic, tabular habit and complex twinning (such as Carlsbad-albite and pericline) of the pristine feldspar. Minor recrystallized constituents may be epidote, garnet, sphene, and granular quartz. *See* METAMORPHIC ROCKS.

Epidiorites have a widespread distribution among low-metamorphic schists and phyllites of mountain ranges in Scotland, the Alps, the Appalachians, and the Rocky Mountains. [T.F.W.B.]

Epidosite

A rare metamorphic rock composed of epidote and quartz. It may form by normal metamorphism of calcareous grits, but most epidosites are of metasomatic-hydrothermal origin. They correspond in a way to skarn rocks, but are formed at a lower temperature. *See* METAMORPHIC ROCKS; METASOMATISM; SKARN.

The original material may be basic lavas or sediments. The truly metasomatic epidosites are found in connection with limestones that have reacted with adjacent silicate rocks under the conditions of the epidote-amphibolite facies. Or they form bands, veins, streaks, or nodules in amphibolites, hornblende schists, and masses of epidiorite. Epidosite from such occurrences is sometimes cut and polished and used for ornamental purposes.

[T.F.W.B.]

Epidote

A mineral belonging to the epidote group, which is a series of complex calcium aluminum sorosilicates. It crystallizes in the monoclinic system; crystals have a prismatic development usually elongated parallel to the *b* axis. Two directions of cleavage yield fragments elongated parallel to the same axis. Hardness is 6½ on Mohs scale; cific gravity is 3.35–3.45. The luster is vitreous the color usually pistachio green to blackish

Well-formed crystals may be black. The composition, $Ca_2(Al,Fe)_3(SiO_4)_3(OH)$, indicates varying amounts of iron and aluminum on which the other properties are dependent. A complete series extends to the iron-free member of the group, clinozoisite. Manganese is present in the pink variety, thulite. Allanite is distinguished from other members of the group by its relatively high content of rare earths. *See* ALLANITE; SILICATE MINERALS.

Epidote is a common mineral in certain metamorphic rocks where it has formed by the alteration of pyroxenes, amphiboles, and biotite. It is characteristic of contact metamorphic deposits in limestone and is there associated with diopside, garnet, and wollastonite. Crystals are found at many localities but the finest are from Knappenwand, Untersulzbachtal, and Salzburg, Austria; and Prince of Wales Island, Alaska.

[C.S.HU.]

Epilepsy

A disease characterized by recurring attacks of loss of consciousness and convulsions. In some cases convulsions may occur without loss of consciousness; in other cases loss of consciousness may occur without convulsions. In symptomatic epilepsy the seizures are the consequences of an organic defect (*see* CONVULSIVE DISORDERS). Epilepsy without demonstrable cause is called idiopathic epilepsy. The disease is frequent; probably one of 200 in the population suffers from it. As a rule the disease starts early in life although it may come at any time.

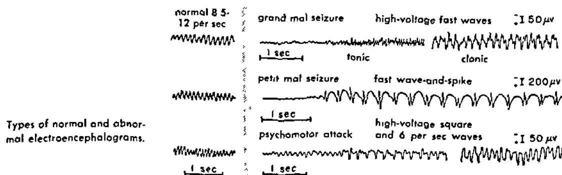
Epileptic types. Convulsive seizures are usually divided into four large groups: (1) Grand mal, or generalized seizures, are the most frequent type. Patients have tonic (without relaxation) seizures, bite their tongues, and become temporarily incontinent. Usually unconsciousness, cyanosis, and tendency to self-injury from falls occur. The patients have amnesia (loss of memory) during the attack and may be confused afterwards. An attack may last for 1–30 min and may be rare or occur frequently, or even in rapid succession, as in status epilepticus. They may occur at any time of the day and are usually preceded by ill-defined sensations called aura, such as weakness, numbness, peculiar odor, and sounds. (2) Local, or Jacksonian, seizures are similar to grand mal, but start usually in one area of the body and then spread to other parts of the body. Often consciousness is retained in such seizures. (3) Petit mal attacks are characterized by brief loss of consciousness, loss of muscular tone, or by contractions (myoclonus) of the muscles. Often the patient briefly interrupts what he was doing. (4) Psychomotor attacks are characterized by longer-lasting loss of consciousness with involuntary, apparently purposeful movements, such as chewing, swallowing, or expectorating, or with movements of the extremities. The attacks may be followed for hours or days by amnesia, mental clouding, and fugue states, during which the patient seemingly is conscious of his behavior, such as traveling and working, but afterward has no conscious recollection of such behavior.

There are no definite signs of epileptic mental symptoms, such as disturbances of intelligence, pathological effects, or psychotic symptoms. Psychotic symptoms and intellectual deficit occur in a very small number of patients; most epileptic patients are mentally normal and competent.

Electroencephalogram. One of the most important findings in epilepsy is electrical tracings (electroencephalogram) from the patient's brain (*see* ELECTROENCEPHALOGRAPHY). The epileptic's electroencephalogram shows that the brain wave rhythm is either abnormally fast or abnormally slow, and usually the slow and fast rhythms alternate. The wave frequency is not constant and there is a tendency to violent and abnormal discharges.

Etiology. The etiology of idiopathic epilepsy is unknown. Undoubtedly it is a disturbance of brain function. The known causes of convulsive seizures, such as trauma and infection, add to a latent predisposition and may produce the manifest syndrome of epilepsy. In any case, hereditary mechanisms play an important role in the etiology of the illness (*see* HUMAN GENETICS). Familial occurrence of clinical and electroencephalographic findings is well known. A thorough search for the various causes of convulsions is important. Idiopathic epilepsy is assumed when no other cause of the illness is found.

Treatment. Once epilepsy was considered a sacred but incurable disease with serious psychiatric and social consequences. Today it is eminently treatable, and epileptics are known to live normally



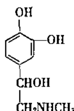
Types of normal and abnormal electroencephalograms.

and usefully. If there is no identifiable cause, the disease must be treated symptomatically. The current treatment with drugs, such as phenobarbital, Dilantin, and Tridione, is very successful. Abstinence from alcohol and avoidance of extreme fatigue are important for the prevention of seizures. Neurosurgical treatment is important only when definite cerebral lesions are found. The organic treatment should be accompanied by the proper psychiatric and social counseling. *See* PSYCHOTHERAPY. [F.C.R.]

Bibliography: F. A. Gibbs and E. L. Gibbs. *Atlas of Electroencephalography*, vol. 2, 1952; H. H. Merritt, *A Textbook of Neurology*, 1955.

Epinephrine

A hormone which is the predominant secretion from the adrenal medulla. It is also known as adrenalin. In 1901, J. Takamine isolated epinephrine in pure form from extracts of adrenal glands, and its structure was finally proved by synthesis by F. Stolz in 1904 as L-(3,4-dihydroxyphenyl)-2-methylaminoethanol:



In mammals, the necessary components for the manufacture of the hormone by the body are the amino acids, tyrosine or phenylalanine. Thus, by labeling these amino acids with radioactive isotopes and tracing them, the rate of formation and normal rate of secretion of adrenal epinephrine can be followed. These rates have been found to be extremely slow, as compared with other hormones, such as the adrenal cortical hormones. This may be explained by the fact that epinephrine is not one of the hormones under the control of the pituitary. *See* PHENYLALANINE; PITUITARY GLAND; TYROSINE.

Epinephrine is a sympathomimetic substance; that is, it acts on tissue supplied by sympathetic nerves, and generally the effects of its action are the same as those of other nerve stimuli. Conversely, the stimulation of the splanchnic or visceral nerves will cause the rapid release of the hormone from the medullary cells of the adrenal gland. Thus, epinephrine plays an important role in preparing the organism to meet conditions of physiologic emergency.

When injected intravenously, epinephrine causes an immediate and pronounced elevation in blood pressure, which is due to the coincident stimulation of the action of the heart and the constriction of peripheral blood vessels. The chief metabolic changes following the injection of epinephrine are a rise in the basal metabolic rate and an increase of blood sugar. Epinephrine effects the latter action by causing an increase in the rate of glycogenolysis or breaking down of stored sugar in the liver. These effects of epinephrine are transitory. *See* ADRENAL GLAND; CARBOHYDRATE; GLYCOGEN. [C.H.L.]

Epithelium

One of the four primary tissues of the body, which constitutes the epidermis and the lining of respiratory, digestive, and genitourinary passages. Its major characteristic is that the cells are close together, separated by a very small amount of intercellular substance. The functions are varied: (1) protective—completely covering the external surface (including the gastrointestinal surface), with the possible exception of the pulmonary alveoli, which may be nearly bare of cells; (2) secretory—secreting fluids and chemical substances necessary for digestion, lubrication, protection, excretion of waste products, reproduction, and the regulation of metabolic processes of the body; (3) absorptive—absorbing nutritive substances and preserving water and salts of the body; (4) sensory—constituting important parts of sense organs, especially of smell and taste; and (5) lubricating—lining all the internal cavities of the body, including the peritoneum, pleura, pericardium, and the tunica vaginalis of the testis.

The epithelium (endothelium) lining the inner cavities of the heart, blood vessels, and lymphatics differs from that of all the other groups because under abnormal conditions it may be transformed into a different type of cell characteristic of connective tissue, the fibroblast. Epithelium may be derived from any of the three primary germ layers of the very early embryo—ectoderm, entoderm, or mesoderm.

The forces which hold the cells together are not satisfactorily understood. The intercellular substance between the cells, also called cement substance, is undoubtedly important. When it is weakened by removing calcium (with versene) or protein (with trypsin or papain), the epithelial cells may be readily separated. The cement substance also contains a carbohydrate moiety. The interdigitation of adjacent cell surfaces and the occurrence of intercellular bridges in certain cells may be important in holding the cells together. Finally, in certain cells, local modifications of contiguous surfaces and the intervening intercellular substances, which together form the terminal bars, may be effective in the same way.

It is not altogether clear what forces prevent the sheet of epithelial cells from falling away from the underlying connective tissue. Undoubtedly the basement membrane of the connective tissue, when it occurs, has a cohesive property. When there is no basement membrane, the epithelium rests directly on the ground substance of the connective tissue, which has similar properties. The area of contact between epithelium and ground substance of connective tissue is frequently increased by irregulari-

ties and processes of the basal surface of the epithelial cells.

The outstanding property of the arrangement of most of the epithelium of the body is the economy of space achieved in the face of a broad exposure of the cell surface. This is especially important at absorptive and secretory surfaces. The efficiency is achieved by the presence of numerous folds, gross and microscopic, temporary or permanent. A part must also be attributed to the surface specialization of the epithelial cells themselves, such as their minute fingerlike processes. Some of these are visible with the light microscope, though many are so small as to be visible only with the electron microscope. Another specialization of the surface or epithelial cell is the occurrence of motile cilia, each of which has a complicated structure. Especially prominent in the cilia are the nine pairs of filaments peripherally disposed in each cilium, with two central ones. The peripheral filaments, and frequently the central ones, join at the base of the cilium to form a basal body which is in turn connected with a rootlet which extends into the cytoplasm and may have a very complicated fine structure.

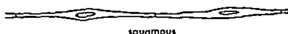
Classification based on morphology. The classification of the varied epithelial tissues is strictly morphological, depending entirely on the shape of the cells and their arrangement. The following is such a classification.

1. Single-layered
   a. Squamous (mesothelium, descending loop of Henle in the kidney)—thin, flat
   b. Cuboidal (ducts, thyroid, choroid plexus)—cubelike
   c. Columnar (intestine), sometimes ciliated (Fallopian tube or oviduct)—tall
2. Multiple-layered or stratified
   a. Squamous (skin, esophagus, vagina)—superficial cells thin and flat, deeper cells cuboidal and columnar
   b. Columnar (pharynx, large ducts of salivary glands), sometimes ciliated (larynx)—two or more layers of tall cells
3. Pseudostratified (male urethra), sometimes ciliated (respiratory passages)—all cells reach to basement membrane, but some extend superficially only part of the way, while others reach the surface
4. Transitional (urinary bladder)—like stratified squamous in the fully distended bladder; in the empty bladder, superficial cells rounded, almost spherical

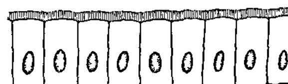
With very few exceptions (stria vascularis of the cochlea, hypertrophied thyroid gland), epithelium is free of blood vessels. Nutrients reach the epithelium, and waste products leave it after passing through the ground substance of adjacent connective tissue.

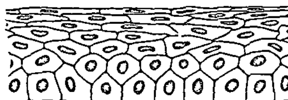
An important property of epithelium is the ability of its cells to glide over surfaces. This allows replacement of dead cells to take place in the normal state, while presenting a closed surface to the external environment; replacement is especially important in wound repair when it is important that a gap in the surface be filled quickly. Gliding ability is also manifested normally in the movement of cells which slide over each other in transitional epithelium when the urinary bladder, for example, is being distended or contracted.

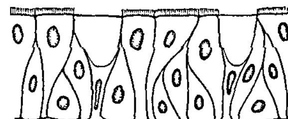
Epithelium of glands. The epithelium of all glands expends energy in forming their secretion. The energy is expended not only in maintaining the integrity of the cells, but also in synthesizing new

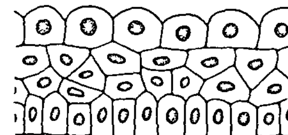

squamous


cuboidal


columnar


stratified squamous


pseudostratified


transitional

Cellular arrangements in epithelial tissues.

substances from the constituents of the blood plasma (as trypsin and insulin by the pancreas) or in changing the concentration of substances already preexisting in the blood plasma (as hydrochloric acid by the stomach, or sweat by the sweat glands of the skin). Glands are classified in various ways. Some commonly used schemes follow.

I. Morphological criteria
  A. Unicellular (mucous goblet cells)
  B. Multicellular
    1. Sheets of gland cells (choroid plexus)
    2. Restricted nests of gland cells (urethral glands)
    3. Invaginations of varying degrees of complexity
      a. Simple or branched tubular (intestinal and gastric glands)—no duct interposed between surface and glandular portion
      b. Simple coiled (sweat gland)—duct interposed between glandular portion and surface
      c. Simple, branched, acinous (sebaceous gland)—glandular portion spherical or ovoid, connected to surface by duct
      d. Compound, tubular glands (gastric cardia, renal tubules)—branched ducts between surface and glandular portion
      e. Compound tubular-acinous glands (pancreas, parotid gland)—branched ducts, terminating in secretory portion which may be tubular or acinar

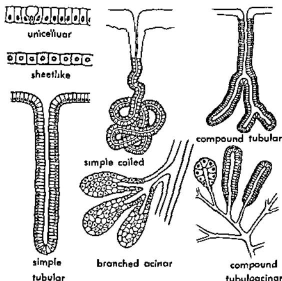
II. Mode of secretion
  A. Exocrine—the secretion is passed directly or by ducts to the exterior surface (sweat glands) or to another surface which is continuous with the external surface (intestinal glands, liver, pancreas, submaxillary gland)
  B. Endocrine—the secretion is passed into the blood stream directly or by way of the lymphatics. These organs are usually circumscribed, highly vascularized, and usually have no connection to an external surface (adrenal, thyroid, parathyroid, islets of Langerhans, parts of the ovary and testis, anterior lobe of the hypophysis, intermediate lobe of the hypophysis, groups of nerve cells of the hypothalamus, and the neural portion of the hypophysis)
  C. Mixed exocrine and endocrine glands (liver, testis, pancreas)
  D. Cytocrine—passage of a secretion from one cell directly to another (melanin granules from melanocytes in the connective tissue of the skin to epithelial cells of the skin)

III. Nature of secretion
  A. Cytogenous (testis, perhaps spleen, lymph node, and bone marrow)—gland "secretes" cells
  B. Acellular (intestinal glands, pancreas, parotid gland)—gland secretes noncellular product

IV. Cytological changes of glandular portion during secretion
  A. Merocrine (sweat glands, choroid plexus) —no loss of cytoplasm
  B. Holocrine (sebaceous glands)—gland cells die and are extruded *in toto* together with the secretory product
  C. Apocrine (mammary gland, axillary sweat gland)—only part of the cytoplasm is extruded with the secretory product

V. Chemical nature of the product
  A. Mucous goblet cells (submaxillary glands, urethral glands)—secretion contains mucin
  B. Serous (parotid gland, pancreas)—secretion does not contain mucin

Glands have been classified according to a variety of criteria. It is not surprising, then, that these categories sometimes overlap. *See* CELL (BIOLOGICAL); GLAND. [I.C.]

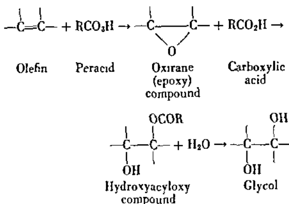


Arrangement of epithelia in various glands.

Epoxidation

A process by which olefinic material may be converted to epoxy (oxirane) compounds. A variety of methods are used. These include direct oxidation in the presence of silver catalysts (*see* ETHYLENE OXIDE), dehydrochlorination of chlorohydrins, and reaction with peracids. The last method is most commonly employed. Depending on the peracid used and reaction conditions, the reaction may proceed beyond the formation of epoxy compounds with the production of hydroxyacyloxy compou[illegible] or glycols.

Epoxides cleave readily with water, a[illegible] cohols, amines, phenols, and mercaptans.



the β-hydroxyethyl derivative is obtained. Under suitable conditions epoxides may be isomerized to ketones. Epoxides are finding uses in insecticides, plasticizers, resins, stabilizers for rubber and paint, lubricants, detergents, emulsifiers, and waxes.

Perbenzoic acid and monoperphthalic acid both give good yields of epoxides when employed under mild conditions in the presence of inert solvents. Since perbenzoic acid can be readily prepared by the oxidation of benzaldehyde with oxygen, a variety of unsaturated materials have been epoxidized by treatment of a solution of the unsaturated material and benzaldehyde with air or oxygen. Monoperphthalic acid is preferred when epoxidation requires extended time for completion. The resulting phthalic acid is insoluble in, and is readily removed from, chloroform solutions.

Peracetic acid is one of the most readily prepared peracids, but early work indicated that it could not be used for epoxidation since the products formed were usually hydroxyacetates or glycols. Later work showed that good yields of oxiranes could be obtained if excess acetic acid is excluded, and an inert solvent is used. A simple method for the epoxidation of a variety of unsaturated materials has now been developed. The reagents used are acetic acid, 30% or higher concentration hydrogen peroxide, and a cationic ion-exchange resin to supply the acid catalyst. The peracetic acid may be prepared externally by passing the reactants through a bed of the resin, or the resin may be added to the combined reaction mixture.

Performic acid, containing excess formic acid, usually gives glycols and hydroxyformates. However, satisfactory yields of oxiranes may be obtained when performic acid is used in a ratio of 0.25–0.5 mole of acid per mole of olefinic compound with at least a full molar proportion of hydrogen peroxide.

The ease with which epoxidation occurs depends on the molecular structure of the unsaturated material, the concentration and ratio of reactants, and the temperature. Mono unsaturated fatty esters such as methyl oleate epoxidize about 10 times

reagents, the more nucleophilic an olefin becomes by replacement of hydrogen with electron-donating groups such as alkyl, the more readily it is epoxidized. Conversely, the more electrophilic an olefin becomes by replacement of hydrogen with electron-attracting groups such as carbonyl or carboxyl groups, the more slowly it is epoxidized. The effect of temperature is illustrated in the epoxidation of methyl oleate with peracetic acid. Using a 20% excess of the peracid, reaction to 80% completion occurs in 1 hour at 45°C, 2 hours at 35°C, and 4 hours at 25°C. *See* ALKENE; SUBSTITUTION REACTION. [C.A.C.]

Bibliography: R. Adams (ed.), *Organic Reactions*, vol. 8, 1954.

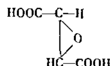
Epoxy resin

A polyether resin formed by the polymerization of bisphenol A and epichlorohydrin. Epoxy resins are used as coatings, adhesives, castings, and foams. Laminates of epoxy resin and glass cloth have been used to make pipes, to repair damaged automobile bodies, and to make small-boat hulls. *See* POLYETHER RESINS.

Bibliography: H. Lee and K. Neville, *Epoxy Resins: Their Applications and Technology*, 1957.

Epoxysuccinic acid

The *trans*-L-isomer of this dicarboxylic acid is produced experimentally by fermentation (*see* FERMENTATION). *trans*-L-*Epoxysuccinic acid* is also known as ethylene oxide-α,β-dicarboxylic acid, and fumaryl-glycidic acid. The compound has a melting point of 180°C and a specific optical rotation of −100° (*see* OPTICAL ACTIVITY). The structural formula is



Selected strains of the filamentous ascomycetous fungus *Aspergillus fumigatus* are inoculated into a medium consisting of commercial grade glucose, 4–10%, and mineral salts. Production can be obtained by either the surface, or tray, process or by the more efficient submerged process, with continuous aeration and agitation. Weight yields of 25% of sugar used have been reported. The acid can be crystallized from vacuum concentrated fermentation liquor. At present, there is no industrial use for this fermentation acid. Various microorganisms convert it in high yield to *meso*-tartaric acid. *See* INDUSTRIAL MICROBIOLOGY. [J.W.F.]

Epsomite

A mineral with the chemical composition $MgSO_4 \cdot 7H_2O$. Epsomite, or epsom salt, occurs in clear, needlelike, orthorhombic crystals. More commonly it is massive or fibrous, although crystals from salt lakes on Kruger Mountain near Orville, Washington, are reported to be several feet long. Fracture

is conchoidal. Luster varies from vitreous to silky. Hardness is 2-2.5 on Mohs scale and specific gravity is 1.68. The mineral has a salty bitter taste and is soluble in water.

Epsomite is found as a capillary coating in limestone caves and in coal or metal mine galleries. It loses readily some of its water of crystallization in dry air. It is also found associated with gypsum and in thin layers in salt deposits of oceanic origin or from salt lakes. *See* MAGNESIUM. [E.C.T.C.]

Equalization, frequency-response

The process of obtaining a desired over-all frequency-response characteristic in an audio-frequency circuit by introducing corrective electrical networks of various types, called equalizers, into the circuit.

Equalizers are used both in communications networks and in systems for the recording and reproducing of photographic film, magnetic tape, and disk phonograph records. These equalizers are in the form of electrical networks of resistance, inductance, and capacitance. The position of the frequency-response characteristics with respect to frequency can be varied by a choice of values of R, L, and C. The maximum variation is 6 db/octave. Larger variations can be obtained by cascading equalizers. Equalizers employing resistance and capacitance are shown in part (a) of the figure. Equalizers employing resistance and inductance are shown in part (b) and equalizers employing resonant circuits consisting of resistance, inductance, and capacitance are shown in part (c).

Referring to the figure, it will be seen that practically any type of equalization can be obtained from the systems and combinations of the systems shown. All major record companies in the United States now use the RIAA (Record Industry Association of America) standard in their recordings; thus the provision of a large number of equalizing or playback settings on hi-fi amplifiers has become unnecessary. *See* AMPLIFIER; DISK RECORDING. [H.F.O.]

Equation of continuity

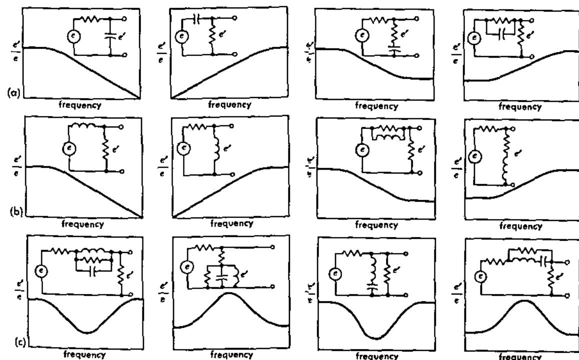
An equation of continuity appears in many branches of physics. In dynamic field theory, it is essentially a statement that charge is conserved or that the rate of increase of charge in any region equals the current **i** flowing into that region. If v is the volume enclosed by the surface S, this statement may be expressed in integral form

$$\int_S \mathbf{i} \cdot \mathbf{n}\, dS = \int_v \nabla \cdot \mathbf{i}\, dv = -\frac{\partial}{\partial t}\int_v \rho\, dv \quad (1)$$

where ρ is the charge density and **n** is a unit vector normal to S. The second integral comes from the first by Gauss' divergence theorem (*see* GAUSS' THEOREM). When currents and charges are confined to a surface S bounded by the curve s, this becomes

$$\oint i \sin\theta\, ds = -\frac{\partial}{\partial t}\int_S \rho\, dS \qquad (2)$$

where θ is the angle between ds and **i** which is di-



Circuit diagrams and frequency-response characteristics of equalizers employing combinations of (a) resistance and capacitance, (b) resistance and inductance, and (c) resistance, inductance, and capacitance.

rected into the area S. This states that the current crossing the boundary of an area equals the rate of increase of charge in the area. The volume in Eq. (1) is arbitrary, so the integrands in the last two integrals are equal. This leads to the differential form

$$\nabla \cdot \mathbf{i} = \frac{\partial i_x}{\partial x} + \frac{\partial i_y}{\partial y} + \frac{\partial i_z}{\partial z} = -\frac{\partial \rho}{\partial t} \qquad (3)$$

where i_x, i_y, and i_z are the rectangular components of $\mathbf{i}$. Maxwell's equations satisfy (3). For application to the motion of charged particles, Eq. (3) would be written

$$\nabla \cdot \rho \mathbf{v} = \rho\left(\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z}\right) = -\frac{\partial \rho}{\partial t} \qquad (4)$$

When $\mathbf{i}$ and ρ vary sinusoidally with time, they may be written as phasors, which are complex numbers such that when multiplied by $e^{j\omega t}$, the real part of the product gives the amplitude, phase, and time dependence. The only change in the preceding equations when $\mathbf{i}$ and ρ are phasor quantities is the replacement of $\partial/\partial t$ by $j\omega$. *See* FLUID-FLOW PRINCIPLES; MAXWELL'S EQUATIONS. [W.R.SM.]

Bibliography: *See* WAVE EQUATION.

Equations, theory of

The branch of mathematics concerned with finding facts concerning the roots of algebraic equations and finding methods for obtaining them. The most important type of algebraic equation is the polynomial equation in one unknown which is an expression of the form $f(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = 0$, where x is called the unknown, or variable, n is a positive whole number, and the a_i, with $i = 0, 1, \ldots, n$, are constants, or fixed numbers, called coefficients of the equation. The left member of the equation is called a polynomial in one variable of degree n. A root of such an equation is a number which when substituted for the variable x makes the left member zero. For example, 3 is a root of the equation $x^3 + 2x^2 - 13x - 6 = 0$. In addition, systems of equations in one or more variables are considered, and here the problem is to find values for the variables which simultaneously satisfy each equation of the system. For a discussion of systems of equations *see* LINEAR SYSTEMS OF EQUATIONS; POLYNOMIAL SYSTEMS OF EQUATIONS.

The topics covered in a systematic study of the theory of equations can be placed in the following principal subdivisions: properties of a polynomial which do not depend on the particular number system containing the coefficients of the polynomial; factorization of polynomials; equations with coefficients which are rational, real, or complex numbers; determination of bounds for real roots, and systematic methods for approximating real roots of equations; the solution of quadratic, cubic, and quartic equations by radicals.

This article is limited to polynomials and to equations which have rational, real, or complex numbers as coefficients. Each of these number systems constitutes a number field.

Let $f(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ and $g(x) = b_mx^m + b_{m-1}x^{m-1} + \cdots + b_1x + b_0 \neq 0$ be polynomials with coefficients in a number field, F. Then the division algorithm, which is a formal statement of the ordinary division process for polynomials, states that there exist unique polynomials $q(x)$ and $r(x)$, with coefficients in F, such that $f(x) = q(x)g(x) + r(x)$, where either $r(x) = 0$, or $r(x)$ has degree less than $g(x)$. If $r(x) = 0$, then $g(x)$ divides $f(x)$. Immediate corollaries of the division algorithm are the remainder and factor theorems. If $g(x)$ is the linear polynomial $(x - a)$, then $f(x) = q(x)(x - a) + r$ where r is a constant. Then, substituting $x = a$, gives $f(a) = r$. This is the remainder theorem. If $f(a) = r = 0$, then a is a root of the equation $f(x) = 0$, giving the factor theorem, which states that a is a root of $f(x) = 0$ if, and only if, $(x - a)$ is a factor of $f(x)$. It follows from the factor theorem that the equation $f(x) = 0$ has at most n roots in F where $f(x)$ is of degree n.

The greatest common divisor (g.c.d.) of two polynomials $f(x)$ and $g(x)$ is a polynomial $d(x)$ which divides both $f(x)$ and $g(x)$, and is divisible by every other polynomial which also divides $f(x)$ and $g(x)$. A process called the euclidean algorithm, based on successive application of the division algorithm, is used to find the g.c.d., $d(x)$, having coefficients in the smallest number field which contains the coefficients of $f(x)$ and $g(x)$. The polynomials $f(x)$ and $g(x)$ are called relatively prime if their greatest common devisor is a constant.

The factorization of a polynomial $f(x)$ depends on the particular number field, F, under consideration. For example, the polynomial $x^5 - \frac{1}{2}x^4 - x^3 + \frac{1}{2}x^2 - 2x + 1$, having coefficients which are rational numbers, factors as $(x^2 + 1)(x^2 - 2)(x - \frac{1}{2})$ over the rational numbers, as

$$(x^2 + 1)(x - \sqrt{2})(x + \sqrt{2})(x - \tfrac{1}{2})$$

over the real numbers, and as

$$(x + i)(x - i)(x - \sqrt{2})(x + \sqrt{2})(x - \tfrac{1}{2})$$

over the complex numbers. A polynomial $f(x)$ with coefficients in F is irreducible over F if it cannot be expressed as a product of polynomials of lower degree. Every polynomial can be expressed in essentially one way as a product of irreducible factors, although there is no general algorithm which enables one to obtain this expression. For polynomials with rational coefficients, there is such an algorithm due to L. Kronecker. There are methods for finding the repeated factors of a polynomial which are often useful as a first step in factoring a polynomial.

If a polynomial equation $f(x) = 0$ with rational coefficients is multiplied by a suitable whole number, an equation with whole number coefficients is obtained. If r/s, a rational number in lowest terms,

is a root of such an equation, then r divides the constant term and s divides the leading coefficient. Hence, the rational roots of $f(x) = 0$ can be found by a finite number of trials.

The fundamental theorem of algebra states that a polynomial equation with complex coefficients has a complex root. From this it follows immediately that a polynomial of degree n with complex coefficients factors into n linear factors over the complex numbers.

If $f(x) = 0$ has real coefficients, then it can be shown that if the complex number $a + bi$ is a root of $f(x)$, the conjugate complex number $a - bi$ is also a root. Thus, the real, irreducible factors of a polynomial with real coefficients are linear or quadratic. A further consequence is that if $f(x)$ with real coefficients has odd degree, then $f(x) = 0$ has at least one real root.

The property of a polynomial $f(x)$ with real coefficients which is basic for the study of the equation $f(x) = 0$ is that $f(x)$ defines a continuous real function. The location principle which follows from this states that if there exist real numbers $a < b$ such that $f(a)$ and $f(b)$ have opposite signs, then $f(x) = 0$ has a real root r such that $a < r < b$. The location principle is used to isolate the real roots and is basic in systematic schemes such as Horner's method and Newton's method for approximating the real roots. A numerical method due to Graeffe can be used to approximate both real and complex roots.

Important results used in finding the real roots of $f(x) = 0$ with real coefficients include the following: Rolles' theorem from differential calculus which states that between two real roots of $f(x) = 0$, there is at least one real root of the derivative $f'(x) = 0$; Sturm's theorem which gives an exact count of the number of real roots in an interval between two real numbers $a < b$; Descartes' rule of signs, which states that the number of positive roots of $f(x) = 0$ equals the number of variations in sign of the coefficients of $f(x)$ minus a non-negative even number. The negative roots of $f(x) = 0$ are positive roots of $f(-x) = 0$.

Results on the bounds for the real roots of $f(x) = 0$ are based on the fact that for sufficiently large values of x, the sign of $f(x)$ is the same as the sign of the leading term $a_n x^n$ of $f(x)$. In particular, $f(x) = 0$ has no real roots for $|x| \geqq |a_k / a_0| + 1$, where a_k is the coefficient of $f(x)$ with the greatest numerical value.

Polynomial equations of degree 2, 3, and 4 are solvable by radicals. This means that there are formulas which give the roots in terms of the coefficients of the equation and these formulas involve only the rational operations and the operation of extraction of roots The principle methods for the solution of the cubic and quartic equations are due to J. Cardan and L. Ferrari respectively. Using the Galois theory of equations, it can be proved that for $n > 4$, there cannot exist a formula involving only rational operations and root extractions for expressing the roots of every polynomial equation of degree n in terms of the coefficients. *See* DETERMINANT; MATRIX THEORY. [R.A.B.]

Bibliography: J. V. Uspensky, *Theory of Equations*, 1948; L. Weisner, *Introduction to the Theory of Equations*, 1938.

Equator

The great circle around the earth, equally distant from the North and South Poles, which divides the earth into Northern and Southern Hemispheres. It is the greatest circumference of the earth because of centrifugal force from rotation, and resultant flattening of the polar areas.

The earth's rotational axis is vertical to the plane of the Equator, and because the inclination of the axis is 66½° (66°55) from the plane of the ecliptic, the plane of the Equator is always inclined 23½° from the ecliptic.

At noon on the days of the vernal and autumnal equinoxes (March 21 and September 23) the sun on the equator is 90° above the horizon. The lowest angle of the sun is 66½°. This occurs at noon, at the winter and summer solstices (June 21 and December 22) when the sun is vertical at the Tropic of Cancer and Tropic of Capricorn respectively. Days and nights are always equal because the plane of the ecliptic intersects the equatorial plane at the earth's center. Consequently, the sun at the equator rises and sets at approximately 6 A.M. and 6 P.M. throughout the year.

The celestial equator, in astronomy, is equally distant from the celestial poles and is the great circle in which the plane of the terrestrial equator intersects the celestial sphere. *See* ASTRONOMY; GEOGRAPHY, MATHEMATICAL. [V.H.E.]

Equilibrium, biological

sidered, and the maintenance of a balanced attitude in a body consisting of a number of parts that are loosely connected by movable joints is a complex matter.

Maintenance of equilibrium. The maintenance of equilibrium is relatively easy in limbless animals, such as the worm, snail, or starfish, that rest and move on an extensive surface of support. The tactile contact between the normal creeping surface and the substrate furnishes sensory information on the animal's orientation with respect to the plumb line. When such an animal is turned on its back, the lack of contact pressure on the creeping surface and the tactile stimulation of the back initiate muscular movements that turn the animal to its normal position again, the righting reflex. The accompanying change in the direction of illumination may also contribute to efficient righting.

Free-swimming and flying animals in their normal upright attitude are often in a precariously poised state of equilibrium. Hence this normal atti-

tude can be maintained only by the continuous operation of corrective equilibrating mechanisms. The same applies to a somewhat lesser degree to long-legged quadrupeds, many mammals, and bipeds, such as birds and some primates including man. In these animals the center of gravity lies some distance above the area of support which, in some cases, such as man, can be relatively small.

Sensory control. The weight of the body must at all times be supported by what are known as antigravity muscles, and the coordination of their activity is controlled by sensory information from a number of sources: (1) the eyes; (2) static and dynamic mechanoreceptors incorporated in the various types of statocysts in invertebrate animals and in the vestibular organ or labyrinth of the vertebrates; (3) proprioceptor organs such as muscle spindles, Golgi endings in tendons, Paccinian corpuscles associated with tendons and joints, and other pressure receptors in supporting surfaces, such as the soles of feet; (4) sensory endings in the viscera, capable of being differentially stimulated by changes in the direction of visceral pull on mesenteries, and other structures.

All these receptor organs supply the central nervous system with information on the orientation of head, trunk, and limbs in the gravitational field.

In daylight, visual fixation of the horizon, or reference to the bright sky as opposed to the ground, can serve as a means of basic orientation This is important in the equilibration of free-swimming and flying animals. Balance in aquatic invertebrates such as shrimps and in fishes is in part contributed to by visual postural mechanisms. These animals turn their backs toward the light, the dorsal light reflex. Some shrimps normally swim belly-up; this shows the ventral light reflex. Among flying animals, insects rely in flight very largely on visual orientation because they are not equipped with special gravity receptors. The compound eye is well suited to the task of keeping the body in a horizontal or near horizontal plane. *See* INSECT PHYSIOLOGY

The subdivision of this eye into separate ommatidia makes it possible to hold the image of the horizon, or other horizontal or vertical landmarks, fixed on a certain part of the eye. If the animal deviates from an even keel, the image leaves the fixation region and this evokes corrective reflex movements designed to return the image to the same part of the eye.

Gravity reception. In many invertebrates and in all vertebrate animals specialized receptor organs exist which are capable of supplying information on the animal's orientation with respect to the plumb line. They are known as statocysts. A statocyst consists of a fluid-filled space, the floor of which is made up of sensory cells and supporting cells. The sensory cells have hair processes which project into the lumen of the statocyst. In the lumen they come into close contact with the statolith, a body of higher specific gravity than the surrounding fluid and tissues. The displacement of the statolith following upon the animal's deviation from an even keel deforms the hair processes and this initiates an impulse response in the sensory nerve. Thus the central nervous system receives information regarding the direction and degree of deviation.

The otolith organs of the vertebrate labyrinth function on the same principle. It has been shown that a given sensory cell in an otolith organ discharges impulses at a frequency which is a function of the position of the animal in space. *See* EAR.

The three semicircular canals of the vertebrate labyrinth, arranged in three different planes in space, are receptor organs for angular accelerations occurring when the moving animal changes direction or when its head is being actively or passively tilted from the normal. Sensory cells located in the fluid-filled ampullae, situated at one end of each canal, are stimulated by the deformation of their hair processes by inertial movements of the fluid. The information derived from a semicircular canal consists in an increase or decrease in the frequency of impulses in the sensory nerve, according to the direction and speed of the rotatory displacement of the head.

Stretch receptors. Stretch receptors, in bilaterally symmetrical limb muscles, register the distribution of body weight. They are parts of a delicately poised feedback control mechanism regulating the tonic state as well as the contraction and relaxation of these muscles, by monitoring their length and tension. In this way they make an important contribution to the maintenance of equilibrium in the stationary and in the moving animal.

In arthropods such receptors are built into the exoskeleton. Receptor hairs, sensitive to prolonged bending, are arranged near the segmental boundaries in trunk and limb joints. They register displacement of these parts relative to each other. Modified hairs, known as campaniform sensilla, respond to strains and stresses in the chitinous armor of insects by the discharge of impulses, thus furnishing information about the distribution of body weight on the limbs. They evoke muscular reflex responses which maintain or restore the normal body posture.

In mammals a complex hierarchy of postural reflexes exists and can be analyzed in experiments based on the progressive elimination of parts of the central nervous system.

If a horse has fallen and is lying on its side, righting is initiated by labyrinthine reflexes and by the asymmetrical tactile stimulation of one flank. These act on the neck muscles and bring about a raising of the head toward its normal position in space. The proprioceptors in the neck muscles are asymmetrically affected by the resulting flexure of the neck, and elicit a series of reflexes which bring first the trunk, then the abdomen, and finally the hindquarters into line with the head. The head is raised and symmetrical vestibular reflexes on the extensor muscles of both front legs stiffen them in

readiness to take the weight of the body when it is transferred on to them by a forward push from the hind limbs. The head is lowered and the hind legs are extended to complete the return of the animal to its normal standing posture.

All these reflexes still occur in mammals after the cerebral hemispheres and the cerebellum have been severed from the brain stem. They depend however on the intactness of the thalamus and midbrain. The removal of the latter brings about decerebrate rigidity characterized by a simultaneous contraction of antigravity muscles. The animal stands rigidly on its extended legs. In such an animal the vestibular reflexes and the proprioceptive reflexes from the neck muscles can be studied. The subsequent destruction of the vestibular organ in the decerebrate animal finally makes possible the demonstration of the isolated spinal proprioceptive components of the righting response [O.L.]

Bibliography: J. F. Danielli and R. Brown (eds.), Physiological mechanisms in animal behaviour, *Symposia Soc. Exptl. Biol.*, 4:60–81, 1950; O. Lowenstein, The equilibrium function of the vertebrate labyrinth, *Biol. Revs.*, 11(1):113–145, 1936.

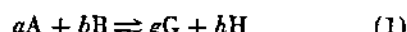
Equilibrium, chemical

In a dynamic or kinetic sense, chemical equilibrium is a condition in which a chemical reaction is occurring at equal rates in its forward and reverse directions, so that the concentrations of the reacting substances do not change with time. In a thermodynamic sense, it is the condition in which there is no tendency for the composition of the system to change; no change can occur in the system without the expenditure of some form of work upon it. From the viewpoint of statistical mechanics, the equilibrium state places the system in a condition of maximum freedom (or minimum restraint) compatible with the energy, volume, and composition of the system. The statistical approach has been merged with thermodynamics into a field called statistical thermodynamics; this merger has been of immense value for its intellectual stimulus, as well as for its practical contributions to the study of equilibria. *See* STATISTICAL MECHANICS; THERMODYNAMICS (CHEMICAL).

Of the three viewpoints, the thermodynamic approach is by far the most powerful and fruitful in treating the quantitative relationships between the position of equilibrium and the factors which govern it. Since thermodynamics is concerned with relationships among observable properties such as temperature, pressure, concentration, heat, and work, the relationships possess general validity, independent of theories of molecular behavior. Because of the simplicity of the concepts involved, this article will utilize that approach.

Chemical potential. Thermodynamics attributes to each chemical substance a property called the chemical potential, which may be thought of as the tendency of the substance to enter into chemical (or physical) change. Although the chemical potential of a substance cannot be directly measured (except on a relative basis), differences in chemical potential are measurable. (The units are those of energy per mole.)

The importance of the chemical potential lies in its relation to what is variously called the affinity or driving force of a chemical reaction Consider the general reaction



$$aA + bB \rightleftharpoons gG + hH \qquad (1)$$

Let μ_A be the chemical potential per mole of substance A, μ_B be the chemical potential per mole of B, and so on. Then, according to one of the fundamental principles of thermodynamics (the second law), the reaction will be spontaneous when the total chemical potential of the reactants is greater than that of the products Thus, for spontaneous change (naturally occurring processes)

$$[g\mu_G + h\mu_H] - [a\mu_A + b\mu_B] < 0 \qquad (2)$$

When equilibrium is reached, the total chemical potentials of products and reactants become equal; thus

$$[g\mu_G + h\mu_H] - [a\mu_A + b\mu_B] = 0 \qquad (3)$$

at equilibrium. The difference in chemical potentials in (2) and (3) is often called the driving force or affinity of the process or reaction; quite naturally, it is zero when the chemical system is in chemical equilibrium.

For reactions at constant temperature and pressure (the usual restraints in a chemical laboratory) the difference in chemical potentials becomes equal to the free energy change, ΔF, for the process

$$\Delta F = [g\mu_G + h\mu_H] - [a\mu_A + b\mu_B] \qquad (4)$$

The decrease in free energy represents the maximum net work obtainable from the process. When no more work is obtainable, the system is at equilibrium. Conversely, if the value of ΔF for a process is positive, some useful work will have to be expended upon the process, or reaction, in order to make it proceed; the process cannot proceed naturally or spontaneously. (The term spontaneously as used here implies only that a process can occur. It does not imply that the reaction will be rapid or instantaneous. Thus, the reaction between hydrogen and oxygen is a spontaneous process in the sense of the term as used here, even though a mixture of hydrogen and oxygen can remain unchanged for years unless ignited or exposed to a catalyst).

Since by definition a catalyst remains unchanged chemically throughout a reaction, its chemical potential does not appear in Eqs. (2), (3), and (4). A catalyst, therefore, can contribute nothing to the driving force of a reaction, nor can it, in consequence, alter the position of the chemical equilibrium in a system. *See* CATALYSIS.

In addition to furnishing a criterion for the equilibrium state of a chemical system, the thermodynamic method goes much farther. In many cases

it yields a relation between the change in chemical potentials (or change in free energy) and the equilibrium concentrations of the substances involved in the reaction. To do this, the chemical potential must be expressed as a function of concentration (and other properties of the substance).

The chemical potential μ is usually represented by the equation,

$$\mu = \mu^0 + RT \ln x + RT \ln f \qquad (5)$$

where R is the ideal gas constant; T is the absolute temperature; $\ln x = \log_e x = 2.3026 \log_{10} x$; x is the concentration of the substance; f is the activity coefficient of the substance; and μ^0 is the chemical potential of the substance in its standard state. See CONCENTRATION SCALES

For substances obeying the laws of ideal solutions (or ideal gases), the last term, $RT \ln f$, is zero, since it is a measure of the deviation from ideal behavior caused by intermolecular or interionic forces. An ideal solution would then be a solution for which $RT \ln f$ is zero over the whole concentration range. For real solutions, the ideal or reference state, where f is unity, is generally chosen as a state of ideal purity (mole fraction = 1) for solids and solvents, and as a state of infinite dispersion (concentration = 0) for gases and solutes. Although the activity coefficient f is regarded as dimensionless, its numerical values will depend upon the particular concentration scale x with which it must be associated

Although the choice of concentration scales is somewhat a matter of convenience, the following are conventionally used: $x = p$ = partial pressure, for gases; $x = c$ or m = molar or molal concentrations, for solutes in electrolytic solutions; x = mole fraction, for solids and solvents. When the choice is not established by convention, the mole fraction scale is to be preferred.

Activity and standard states. It is often convenient to utilize the product, fx, called the activity of the substance and defined by $a = fx$. The activity may be looked upon as an effective concentration of the substance, measured in the same units as the concentration x with which it is associated. The standard state of the substance is then defined as the state of unit activity (where $a = 1$), and is characterized by the standard chemical potential, μ^0. Clearly, the terms μ^0, f, and x are not independent; the choice of the activity scale serves to fix the standard state. For example, for an aqueous solution of hydrochloric acid, the standard state for the solute (HCl) would be an (hypothetical) ideal 1 molar (or molal) solution, and for the solvent (H_2O) the standard state would be pure water (mole fraction = 1). The reference state would be an infinitely dilute solution; here the activity coefficients would be unity for both solute and solvent. For the vapor of HCl above the solution, the standard state would be the ideal gaseous state at 1 atm partial pressure; the reference state would be a state of zero pressure. (For gases, the term fugacity is used instead of activity.)

It should be noted that the reference state is a limiting state which in many cases can be reached only through an extrapolation from observed behavior.

Equilibrium constant. If the general reaction (1) occurs at constant temperature T and pressure P when all of the substances involved are in their standard states of unit activity, Eq. (4) would become

$$\Delta F^0 = [g\mu^0_G + h\mu^0_H] - [a\mu^0_A + b\mu^0_B] \qquad (6)$$

The quantity ΔF^0 is known as the standard free energy change for the reaction at that temperature and pressure for the chosen standard states. (Standard state properties are commonly designated by a superscript, ΔF^0, μ^0.) Since each of the standard chemical potentials (μ^0) is a unique property determined by the temperature, pressure, standard state, and chemical identity of the substance concerned, the standard free energy change ΔF^0 is a constant (parameter) characteristic of the particular reaction for the chosen temperature, pressure, and standard states.

If, in a reaction (1) at constant temperature and pressure, the chemical potential of each substance is expressed in terms of Eq. (5), the free energy change for the reaction, from (4), becomes

$$\Delta F = \Delta F^0 + RT \ln \frac{a_G{}^g a_H{}^h}{a_A{}^a a_B{}^b} \qquad (7)$$

in terms of the activities and the standard free energy change (6). Equation (7) is often written in the form

$$\Delta F = \Delta F^0 + RT \ln Q^0 \qquad (8)$$

where Q^0 is the ratio of the activities of products to the activities of reactants, each activity bearing as an exponent the corresponding coefficient in the balanced equation for the reaction. The standard free energy change ΔF^0 serves as a reference point from which the actual free energy change ΔF can be calculated in terms of the activities of the reacting substances.

When the system has come to chemical equilibrium at constant temperature and pressure, $\Delta F = 0$, from Eq. (3). Equation (7) then leads to the very important relation

$$\Delta F^0 = -RT \ln K^0 \qquad (9)$$

where

$$K^0 = \left[\frac{a_G{}^g a_H{}^h}{a_A{}^a a_B{}^b}\right] \qquad (10)$$

and the activities are the equilibrium values. The ratio of the activities at equilibrium, K^0, is called the equilibrium constant or more precisely, the thermodynamic equilibrium constant. (The terms K^0 and Q^0 denote the ratios of activities; this usage is currently developing.) The equilibrium constant is a true constant, characteristic of the reaction, since it is determined uniquely in terms

of the standard free energy change. The term $RT \ln K^0$ represents the maximum net work which the reaction could make available when carried out at constant temperature and pressure with all its substances in their standard states.

Equations (8) and (10) can be combined in the form

$$\Delta F = RT \ln \frac{Q^0}{K^0} \qquad (11)$$

When Q^0 for a specified set of conditions is larger than K^0 (so that ΔF is positive), the proposed reaction can not occur. On the other hand, when Q^0 is less than K^0, the proposed process or reaction can occur. The equilibrium constant thus serves as a measure of the position of chemical equilibrium for a system. For the proposed process for which Q^0 is greater than K^0, the reaction system would be moving away from its equilibrium state (impossible of its own accord!), and for a proposed process for which Q^0 is less than K^0, the process would bring the system closer to its equilibrium state (as in all naturally occurring processes).

Instead of activities, values of concentrations and activity coefficients at equilibrium may be used to express the equilibrium constant,

$$K^0 = \left[\frac{x_G{}^g x_H{}^h}{x_A{}^a x_B{}^b}\right] \left[\frac{f_G{}^g f_H{}^h}{f_A{}^a f_B{}^b}\right] \qquad (12)$$

This gives the equilibrium constant as a product of two terms, each of the same form as K^0 itself. The first term, involving concentrations, is directly measurable if the system can be analyzed at equilibrium. On the other hand, the activity coefficient term,

$$\Gamma = \frac{f_G{}^g f_H{}^h}{f_A{}^a f_B{}^b} \qquad (13)$$

commonly is difficult to evaluate.

The intense study of activity coefficients which this evaluation difficulty engendered has led to a number of simplifying principles and some useful theoretical treatments of the subject. For gases, the activity coefficients differ only slightly from unity for pressures up to 10 atm and can be evaluated from equation-of-state data. For mixtures of nonelectrolytes, the values also appear to be close to unity in many cases. For solutions of electrolytes, the activity coefficients vary greatly with concentration, and in many cases approach unity only below a useful or even meaningful concentration. The theoretical treatments of P. Debye, E. Hückel, and others have systematized the patterns of electrolyte behavior, making possible a reasonable estimate of the activity coefficients in many cases. Many tables of experimental data are available for electrolytes. *See* ACTIVITY (THERMODYNAMICS); FUGACITY; SOLUTION.

For many approximate calculations or when data for Γ are scarce, it is common to express the equilibrium constant as the concentration term only.

That is,

$$K = \frac{x_G{}^g x_H{}^h}{x_A{}^a x_B{}^b} \qquad (14)$$

Unless Γ is a rather insensitive function of concentration, the so-called constants obtained in this manner will not be constant at all as the composition is varied, and even though approximately constant, may vary considerably from the true value of K^0. Although the practice of assigning Γ a value of unity will often give adequate results and is frequently the only expedient available, the results should be used with caution.

It is appropriate to point out here that the kinetic concept of chemical equilibrium introduced by C. M. Guldberg and P. Waage (1864) led to the formulation of the equilibrium constant in terms of concentrations. Although the concept is correct in terms of the dynamic picture of opposing reactions occurring at equal speeds, it has not been successful in coping with the problems of activity coefficients and cannot lead to the useful relations (4 to 11) in terms of an energetic criterion for the position of equilibrium. Conversely, the thermodynamic approach yields no relationship between the driving force of the reaction and the rate of approach to equilibrium. *See* KINETICS (CHEMICAL).

The influence of temperature upon the chemical potentials, and hence upon the equilibrium constant, is given by the Gibbs-Helmholtz equation

$$\left[\frac{d \ln K^0}{dT}\right]_P = \frac{\Delta H^0}{RT^2} \qquad (15)$$

The derivative on the left represents the slope of the curve obtained when values of $\ln K^0$ for a reaction, obtained at different temperatures but always at the same pressure P, are plotted against temperature. The standard heat of reaction ΔH^0 for the temperature T at which the slope is measured, is the heat effect which could also be observed by carrying out the reaction involving the standard states in a calorimeter at the corresponding temperature and pressure.

For endothermic reactions, which absorb heat (ΔH^0 positive), K^0 increases with increasing temperature. For exothermic reactions, which evolve heat (ΔH^0 negative), K^0 decreases with increasing temperature and the yield of products is reduced. A more useful arrangement of Eq. (15) is

$$\left[\frac{d \ln K^0}{d(1/T)}\right]_P = \frac{-\Delta H^0}{R} \qquad (16)$$

In practice, plots of $\ln K^0$ against $1/T$ are nearly linear for many reactions where the value of ΔH^0 changes slowly with temperature. Hence, over small temperature ranges, Eq. (16) becomes, in integrated form,

$$\ln \frac{K_2{}^0}{K_1{}^0} = \frac{\Delta H^0}{R} \left[\frac{T_2 - T_1}{T_1 T_2}\right] \qquad (17)$$

This relation is much used for calculating heats reaction from two equilibrium measu-

for determining a new equilibrium constant K_2^0 from values of K_1^0 and ΔH^0.

For accurate work, or for extending the calculations over a wide range of temperature, ΔH^0 must be known as a function of temperature before Eq. (15) or (16) can be integrated. When sufficient heat capacity data are available, Kirchhoff's equation

$$\Delta H^0 = \Delta H_0^0 + \alpha T + \frac{\beta T^2}{2} + \frac{\gamma T^3}{3} \qquad (18)$$

may be combined with Eq. (15) to yield

$$R \ln K^0 = -\frac{\Delta H_0^0}{T} + \alpha \ln T + \frac{\beta T}{2} + \frac{\gamma T^2}{6} + I \qquad (19)$$

In these equations, α, β, and γ are determined from heat capacity data; the constants ΔH_0^0 and I require knowledge of one value of ΔH^0 and one value of K^0, or values of K^0 at two temperatures. *See* HEAT CAPACITY; THERMOCHEMISTRY.

When equilibria are studied under conditions of constant temperature and constant volume, and using volume concentrations to fix standard states, the preceding treatment will yield ΔE^0, the internal energy change, which is the calorimetric heat of reaction at constant volume.

Homogeneous equilibria. These involve single-phase systems: gaseous, liquid, and solid solutions. In most cases, solid solutions are so far from ideal that equilibrium constants are impractical and are better treated in terms of the phase rule (*see* EQUILIBRIUM, PHASE).

A typical gas-phase equilibrium is the ammonia synthesis:



$$N_2 + 3H_2 \rightleftharpoons 2NH_3 \qquad (20)$$

The most natural concentration measures are mole fraction or partial pressure; molar concentrations might be used. The partial pressures p_i are defined in terms of mole fraction N_i, and the total pressure P, by $p_i = N_iP$; note that partial pressures in general are not directly observable. For low pressures, where activity coefficients are practically unity,

$$K_p^0 = \frac{p_{NH_3}^2}{p_{N_2}p_{H_2}^3} = \frac{N_{NH_3}^2}{N_{N_2}N_{H_2}^3}P^{-2} \qquad (21)$$

or

$$K_p^0 = K_N^0 P^{\Delta n} \qquad (22)$$

where Δn is the increase in the number of moles of gases (here $\Delta n = -2$). The mole fraction equilibrium constant K_N^0 is pressure dependent, but K_p^0 is independent of pressure because of the difference in standard states. Consequently, an increase in total pressure P must lead to an increase in K_N^0 in this case. If the increase in total pressure is due to a decrease in volume of the system, the result will be an increased yield of products (NH_3). An increase in pressure brought about by injection of an inert gas into a constant volume system would not affect the partial pressures of the reacting gases nor the ultimate yield of products. The value of K_N^0, and thus that of ΔF_N^0, will be affected by change in total pressure due to a change in the net work of mixing and unmixing the gases. This reaction (20) is exothermic, so best yields will be obtained at lower temperatures (see Eq. 15).

A typical liquid-phase equilibrium is the dissociation of acetic acid in water,



$$HC_2H_3O_2 + H_2O \rightleftharpoons H_3O^+ + C_2H_3O_2^- \qquad (23)$$

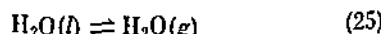
$$\frac{K^0}{\Gamma} = \frac{m_{H_3O^+}m_{C_2H_3O_2^-}}{m_{HC_2H_3O_2}N_{H_2O}} \qquad (24)$$

In this particular case, the activity coefficient ratio Γ is slightly less than unity in dilute solutions; the reference state will be the infinitely dilute solution. The mole fraction of water in the applicable concentration range is close to unity, and so numerically plays little part in evaluation of K^0. For this reason, it is commonly omitted in the formulation of the equilibrium constant; many erroneous explanations exist in the literature. Letting α be the fraction of acetic acid dissociated and m be the total concentration, $K^0/\Gamma = m\alpha^2/(1-\alpha)$; inspection shows that α increases with dilution.

The solvent appears to be inert, since its chemical potential remains practically unchanged over the useful concentration range. As a result of this apparent inertness of the solvent, it is not possible to determine the extent of hydration of any dissolved species from equilibrium studies. Thus whether the actual ion is H^+, H_3O^+, or $H_5O_2^+$, it is the total stoichiometric concentration that is measured and used in Eq. (24). *See* EQUILIBRIUM, IONIC.

Heterogeneous equilibria. These are usually studied at constant pressure, since at least one of the phases will be a solid or liquid. The imposed pressure may be that of an equilibrium gaseous phase, or it may be an externally controlled pressure.

In the equilibrium,



$$H_2O(l) \rightleftharpoons H_2O(g) \qquad (25)$$

$$\Delta F^0 = \mu_g^0 - \mu_l^0 = -RT \ln \frac{p}{N} \qquad (26)$$

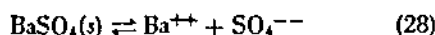
Here $K^0 = p/N$, the ratio of the vapor pressure p to the liquid mole fraction N. For pure water, the equilibrium constant is simply the standard vapor pressure p^0, and the Clausius-Clapeyron equation is just a special case of the Gibbs-Helmholtz equation (15). Now when a small amount of solute is added, decreasing the mole fraction of solvent, the vapor pressure p must be lowered to maintain equilibrium (Raoult's law). The effect of the total applied pressure P upon the vapor pressure p of the liquid is given by the Gibbs-Poynting equation

$$\left[\frac{dp}{dP}\right]_T = \frac{V_l}{V_g} \qquad (27)$$

where V_l and V_g are the molar volumes of liquid and vapor. The vapor pressure will increase as external pressure is applied (activity increases with pressure). If the external pressure is applied to a

solution by a semipermeable membrane, an applied pressure *can be found which will restore the vapor* pressure (or activity) of the solvent to its standard state value. *See* OSMOSIS.

Solubility equilibria are of wide variety. For a solid, such as barium sulfate, $BaSO_4$,

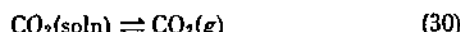


$$BaSO_4(s) \rightleftharpoons Ba^{++} + SO_4^{--} \qquad (28)$$

$$K^0 = \frac{a_{Ba^{++}} a_{SO_4^{--}}}{a_{BaSO_4}} \qquad (29)$$

When the solid state is pure, its mole fraction will be unity. If it is extremely finely divided, its activity coefficient will become greater than unity; with this increase in the activity of the solid state, the solubility must increase to maintain equilibrium. On the other hand, inclusion of foreign ions in the crystal lattice (solid solution formation) may lower the activity of the solid state.

For a gas dissolved in a liquid,



$$CO_2(soln) \rightleftharpoons CO_2(g) \qquad (30)$$

$$K^0 = \frac{p}{N} \qquad \text{(Henry's law)} \qquad (31)$$

When the gas dissociates in the liquid,



$$HCl(aqueous) \rightleftharpoons HCl(g) \qquad (32)$$

$$K^0 = \frac{p}{m^2} \quad \Gamma \qquad (33)$$

where m is the molal concentration, the equilibrium constant must reflect this behavior through the proper exponents. Equation (33) might correctly imply that the molecules were dimerized in the gaseous phase, if it were not known from a study of vapor densities that in this case they are not.

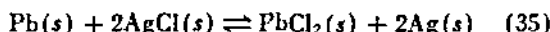
Similarly, when a solute distributes itself between two immiscible phases, the equilibrium constant takes the form

$$K^0 = \frac{c_1}{c_2^n} \quad \Gamma \qquad (34)$$

where n is ratio of the molecular weight in phase 1 to that in phase 2. The equilibrium concentrations c in Eq. (34) reflect the relative solubilities of the solute in the two phases; since solubilities may vary widely, distribution operations may provide an effective means for concentrating a widely dispersed solute. *See* EXTRACTION.

When two components form two immiscible phases at equilibrium, each condensed phase will be a *saturated solution; complete immiscibility* is impossible in principle, since the chemical potential of any component must be the same in both phases. The separation of a liquid system into two liquid phases is a manifestation of the nonideality of the solutions. For practical purposes, however, many solids may be regarded as immiscible because of the stringent requirements associated with formation of the crystal lattice.

In reactions involving condensed and immiscible phases, for example,



$$Pb(s) + 2AgCl(s) \rightleftharpoons PbCl_2(s) + 2Ag(s) \qquad (35)$$

there can be no change in concentration of any phase during the reaction. Then the Q^0 term in Eq (10) will be constant, and ΔF can never become zero. Although reaction is possible, there can be no equilibrium until one of the reactants is completely used up. There might, of course, be one condition of temperature and pressure for which ΔF could be zero. Such is the case in transition phenomena, or melting-freezing phenomena. From the phase-rule viewpoint, the system in Eq. (35) lacks one degree of freedom if reaction is possible, so one phase must disappear in order to attain equilibrium. [C.E.A.]

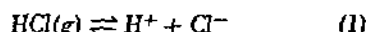
Bibliography: J. G. Aston and J. J. Fritz, *Thermodynamics and Statistical Thermodynamics*, 1959; F. Daniels and R. A. Alberty, *Physical Chemistry*, 1955; S. Glasstone, *Textbook of Physical Chemistry*, 2d ed., 1946.

Equilibrium, ionic

An equilibrium in a chemical reaction in which at least one ionic species is produced, consumed, or changed from one medium to another.

Types of equilibrium. A few examples can illustrate the wide variety of types of ionic equilibrium which are known.

Dissolution of an unionized substance. The dissolution of hydrogen chloride (a gas) in water (an ionizing solvent) can be used to illustrate this type.



$$HCl(g) \rightleftharpoons H^+ + Cl^- \qquad (1)$$



$$HCl(g) + H_2O \rightleftharpoons H_3O^+ + Cl^- \qquad (2)$$



$$HCl(g) + 4H_2O \rightleftharpoons H_9O_4^+ + Cl^- \qquad (3)$$

Equations (1), (2), and (3) represent exactly the same equilibrium. Equation (1) ignores the hydration of the proton and is preferred for many purposes when the hydration (or solvation) of the proton is irrelevant to a particular discussion. Equation (2) is written in recognition of the widely held belief that free protons do not exist in aqueous solution. Equation (3) represents very recent work which indicates that another three molecules of water are very firmly bound to the H_3O^+ ion (the oxonium ion). There is no implication in Eq. (3), however, that the total number of molecules of water attached to, or weakly affected by, the oxonium ion may not be considerably larger than three.

Not much is known about the solvation of ions, although it has been proved that each chromic ion, Cr^{3+}, in dilute aqueous solution holds at least six water molecules. Only a few other similar data have been clearly established. Hence equations for the other examples cited below are written without regard to solvation except that the hydrogen ion is usually written in accordance with present common practice as H_3O^+. *See* ACID AND BASE.

Dissolution of a crystal in water. The dissociation of solid silver chloride illustrates this type of equilibrium.



$$AgCl\ (crystal) \rightleftharpoons Ag^+ + Cl^- \qquad (4)$$

See SOLUBILITY PRODUCT CONSTANT.

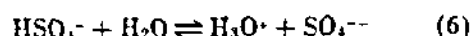
Dissociation of a strong acid. Nitric acid, HNO_3, dissociates as it dissolves in water.



$$HNO_3 + H_2O \rightleftharpoons H_3O^+ + NO_3^- \qquad (5)$$

At 25°C about one-half the acid is dissociated in a solution containing 10 (stoichiometric) moles of nitric acid per liter.

Dissociation of an ion in water. The bisulfate ion, HSO_4^-, dissociates in water.



$$HSO_4^- + H_2O \rightleftharpoons H_3O^+ + SO_4^{--} \qquad (6)$$

About one-half the HSO_4^- is dissociated in an aqueous solution containing about 0.011 mole of sulfuric acid per liter at 25°C.

Dissociation of water itself. In pure water at 25°C the concentration of each ion is about 10^{-7} mole/liter, but increases rapidly as temperature is increased.



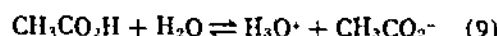
$$2H_2O \rightleftharpoons H_3O^+ + OH^- \qquad (7)$$

Formation of a complex ion. In water or in a mixture of fused (chloride) salts, complex ions such as $ZnCl_4^{--}$ may be formed. *See* COMPLEX COMPOUNDS.



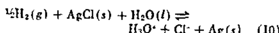
$$Zn^{++} + 4Cl^- \rightleftharpoons ZnCl_4^{--} \qquad (8)$$

Dissociation of a weak acid. In water, acetic acid dissociates to form hydrogen (hydronium) ion and acetate ion.



$$CH_3CO_2H + H_2O \rightleftharpoons H_3O^+ + CH_3CO_2^- \qquad (9)$$

Electrochemical reaction. Reaction (10) takes place "almost reversibly" when



$$\tfrac{1}{2}H_2(g) + AgCl(s) + H_2O(l) \rightleftharpoons H_3O^+ + Cl^- + Ag(s) \qquad (10)$$

A small current is allowed to flow through an electric cell consisting of an aqueous solution of HCl saturated with silver chloride, a hydrogen electrode, and a silver electrode. Saturation is maintained by an excess of solid silver chloride which for convenience is sometimes mixed with the silver or plated, as a coating, on the metal. The electrode is then called a silver-silver chloride electrode.

Many more types of equilibria could be mentioned, including reactions occurring entirely in the gaseous phase and reactions occurring between substances dissolved in two immiscible liquids.

Quantitative relationships. Each reaction obeys an equilibrium equation of the following type:

$$\frac{[H^+][Cl^-]}{[HCl(g)]}\,\frac{f_+f_-}{\gamma_g} = Q_cQ_f = K \qquad (11)$$

The activity coefficient γ_g can be ignored here because it is very nearly unity. The terms f_+ and f_- are the respective activity coefficients of H^+ and Cl^- but cannot be determined separately. Their product can be determined experimentally and can also be calculated theoretically for very dilute solutions by means of the Debye-Hückel theory of interionic attraction. Because γ_g and f_+f_- are nearly unity, Eq. (10) demands that the pressure of HCl gas above a dilute aqueous solution be proportional to the square of the concentration of the solute. Q_c is called the concentration quotient and Q_f the quotient of activity coefficients.

Similarly the dissociation of acetic acid obeys the equation

$$\frac{[H^+][CH_3CO_2^-]}{[CH_3CO_2H]}\,\frac{f_+f_-}{f_u} = Q_cQ_f = K \qquad (12)$$

Early work on electrolytes revealed that Q_c, the concentration quotient, was constant within the limits of accuracy attainable at the time. More recent work has revealed that the measured concentration quotient Q_c first increases as concentration is increased from very small values and then decreases sharply. The initial increase is due largely to the electrical forces between the ions which reduce the product f_+f_-. There is some evidence that the subsequent decrease in Q_c and the concomitant rise in Q_f are due to removal of some of the monomeric acetic acid by the formation of dimeric acetic acid, $(CH_3CO_2H)_2$. The fact is, however, that knowledge concerning activity coefficients in solutions other than very dilute ones is not yet understood. Even the experimental methods for the measurement of the molecular species involved in some equilibria were not evolved until recently. *See* ELECTROLYTIC CONDUCTANCE; HYDROLYSIS.

[T.F.Y.]

Bibliography: J. Bjerrum et al., Stability constants of metal-ion complexes, with solubility products of inorganic substances, *Chem. Soc. London*, pts. 1 and 2, 1957–1958; D. E. Gray (ed.), *American Institute of Physics Handbook*, 1957; W. M. Latimer, *Oxidation Potentials*, 2d ed., 1952.

Equilibrium, phase

A general field of physical chemistry dealing with the various situations in which two or more phases (or states of aggregation) can coexist in thermodynamic equilibrium with each other, with the nature of the transitions between phases, and with the effects of temperature and pressure upon these equilibria. Many superficial aspects of the subject are largely qualitative, as for example the empirical classification of types of phase diagrams; but the basic problems always are susceptible to quantitative thermodynamic treatment, and in many cases, statistical thermodynamic methods can be applied to simple molecular models.

Thermodynamics requires that when two phases, α and β, are free to exchange heat, mechanical work, and matter (chemical species), the temperature T, the pressure P, and the chemical potential (partial molar free energy) μ_i of each particular component i must be equal in both phases at equilibrium.

$$T_\alpha = T_\beta \qquad P_\alpha = P_\beta \qquad \mu_{i,\alpha} = \mu_{i,\beta} \qquad \mu_{j,\alpha} = \mu_{j,\beta}$$

These conditions of thermal, mechanical, and material equilibrium need not all be present if the equilibrium between phases is subject to inhibiting restrictions. Thus, for a solution of a nonvolatile solute in equilibrium with the solvent vapor, the condition of equality of solute chemical potentials $\mu_{2,a} = \mu_{2,\beta}$ need not apply, since there can be no solute molecules in the vapor phase. Similarly, in osmotic equilibria, in which solvent molecules can pass through a semipermeable membrane, whereas solute molecules cannot, $\mu_{1,a} = \mu_{1,\beta}$ and $T_{1,a} = T_{2,\beta}$, but the solute chemical potentials μ_2 are unequal, as are the pressures on opposite sides of the membrane. *See* OSMOSIS; SOLUTION.

If a system consists of P phases and C distinguishable components, there are $C+2$ thermodynamic variables (C chemical potentials μ_i, plus the temperature and pressure) which are interrelated by an equation for each phase. Since there are P independent equations relating the $C+2$ variables, one needs to fix only $F = C + 2 - P$ variables to define completely the state of the system at equilibrium; the other variables are then beyond control. This relation for the number of degrees of freedom F, or variance, is called the Phase Rule and was first derived by Willard Gibbs in 1873. It has proved to be a powerful tool in interpreting and classifying types of phase equilibria.

When chemical changes may occur in the system, the number of components C is the number of independent components, that is, the number of components whose amounts can be varied by the experimenter; this is equal to the total number of chemical species present less the number of independent chemical equilibria between them.

An invariant system has no degrees of freedom ($F = 0$), for which the number of phases $P = C + 2$. For a one-component system, such an invariant point is a triple point at which three phases coexist at a single temperature and pressure only; for a two-component system, a quadruple point (four phases) would be invariant. *See* TRIPLE POINT

In a univariant system ($F = 1$), $P = C + 1$. With a one-component system, one can fix the temperature at which two phases (liquid and gas, for instance) can coexist in equilibrium, then the pressure (here the vapor pressure) is determined, and not subject to external control. A univariant system is described by a line in a phase diagram, for example, a plot of vapor pressure versus temperature. The differential equation for such a univariant line was first deduced by B. P. E. Clapeyron (1834):

$$\frac{dP}{dT} = \frac{\Delta H}{T\Delta V}$$

where ΔH and ΔV are the enthalpy change (heat absorbed) and volume change, respectively, for the transition from one phase to another. *See* VAPOR PRESSURE.

The Clapeyron equation, although most extensively applied to one-component systems, is applicable in fact to all univariant systems. For example, in the chemical decomposition of solid calcium carbonate to solid calcium oxide and gaseous carbon dioxide



$$CaCO_3(s) \rightleftharpoons CaO(s) + CO_2(g)$$

the three-phase system (two solids and one gas) is univariant (since the number of independent components C is two), the equilibrium pressure is the decomposition pressure of $CaCO_3$, and the ΔH and ΔV of the Clapeyron equation are the enthalpy and volume changes associated with the chemical reaction.

BINARY SYSTEMS

Phase diagrams of binary systems containing two components are easily classified. Typical examples of the important classes (liquid-gas, liquid-liquid, solid-liquid, solid-solid) have been selected for description.

Liquid-gas equilibrium. In a one-component system, liquid and vapor are in equilibrium at the boiling point. For two-component systems, the two-phase situation is bivariant and more complex. A complete temperature-pressure-composition diagram would be three-dimensional; so most phase diagrams are made for either constant pressure or constant temperature. Figure 1 shows the simplest type of binary liquid-vapor temperature-composition diagram, exemplified by the system carbon tetrachloride–stannic chloride which forms essentially ideal solutions. The regions labeled G and L are one-phase regions, gas (vapor) and liquid respectively; the region labeled $L + G$ is a two-phase region in which liquid and vapor coexist. If

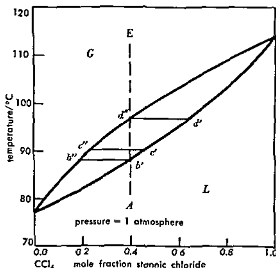


Fig. 1. Temperature-composition diagram for the system carbon tetrachloride + stannic chloride. G = gas (vapor) phase; L = liquid phase; $L + G$ = two coexisting phases, liquid and gas.

the temperature of a liquid mixture of 40 mole % $SnCl_4$ (mole fraction = 0.40) is increased at a constant pressure of 1 atm, the change in the system can be traced along the straight line $ab'cd''e$. At low temperatures, only one phase, the liquid, is present, but at 87.5°C (point b'), a vapor phase appears. The composition of this vapor phase is given by point b'' (mole fraction $SnCl_4$ = 0.18), and the two conjugate phases are connected on the diagram by the tie line $b''b$. As the temperature is increased further, more vapor is formed; since the vapor is rich in CCl_4, this component becomes relatively depleted in the liquid phase, and the liquid composition moves along the line $b'c'd'$, while the vapor composition moves along the line $b''c''d''$

At 90°C, the over-all composition of the two-phase system is represented by point c, but the compositions of vapor and liquid separately are given by the two ends of the tie line, points c'' and c', respectively (mole fractions of 0.22 and 0.47, respectively). The relative amounts of the two phases are given by the lever-arm principle of physics. The ratio of the moles of vapor to moles of liquid is given by the ratio of the length cc' to the length $c''c$, here 0.07/0.18, or 28% in the vapor phase. Further increase in temperature produces more and more vapor until, at 97°C, the liquid phase (point d', mole fraction 0.62) has become vanishingly small; at higher temperatures, it disappears, and only the vapor phase (point d'', mole fraction = 0.40) remains. Further increase in temperature (along the line $d''e$) is uneventful.

In this simple system, there are no maxima or minima in the liquid and vapor curves; consequently, such liquid mixtures can be separated completely into the two pure components by fractional distillation. Systems with maximum boiling points (acetone + chloroform, Fig. 2) or minimum boiling points (ethanol + benzene, Fig. 3) cannot be so separated into the pure substances. At the maximum or minimum, the composition of the liquid is identical with that of the vapor with which it is in equilibrium; continued boiling will not alter these compositions, so these solutions are called constant-boiling mixtures or azeotropes. It should be noted that the solid and liquid lines are smooth curves tangent to each other at such a point; any phase diagram which shows a sharp corner is thermodynamically incorrect. *See* AZEOTROPIC MIXTURE; DISTILLATION.

Maximum boiling mixtures are associated with negative deviations from ideal behavior; this type of deviation usually arises from strong attractions between molecules of the different species (sometimes called compound formation). Minimum boiling mixtures are associated with positive deviations from ideal behavior; this usually arises when the attraction between two unlike molecules (1—2) is weaker than the average of two like pairs (1—1 and 2—2); extreme examples of this arise when one component may be described as associated. The simpler type of phase diagram (Fig. 1) occurs when the two components mix nearly ideally or when the boiling points are very different. *See* MOLECULAR ASSOCIATION.

Liquid-liquid phase equilibrium. When two liquids are sufficiently different in their intermolecular forces, they may not mix in all proportions, but instead be only partially miscible, part of the phase diagram being occupied by a two-phase region of two immiscible liquid phases. Most liquids (but not all) become more miscible as the temperature increases and are completely miscible at the critical solution temperature (also called the consolute temperature). The system aniline + n-hexane (Fig. 4) is a typical example of such liquid-liquid immiscibility and critical phenomena. The liquid-

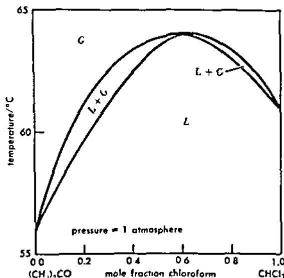


Fig. 2. Temperature-composition diagram for acetone + chloroform, showing maximum boiling point.

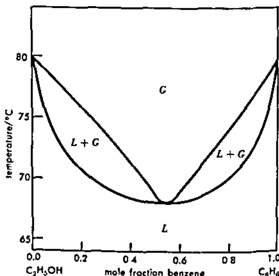


Fig. 3. Temperature-composition diagram for ethanol + benzene, showing minimum boiling point.

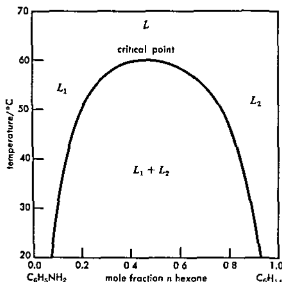


Fig. 4. Liquid liquid equilibrium in the system aniline + n-hexane.

liquid phase boundary and the critical solution temperature are only slightly dependent upon pressure. The size of the two-phase region increases with pressure if (as is usual) the two liquids expand on mixing at constant pressure.

Solid-liquid phase equilibria. When two substances are completely miscible with each other and form a complete series of solid solutions, the solid-liquid phase diagrams are entirely analogous to the liquid-vapor diagrams illustrated above. Those with no maximum or minimum, usually associated with nearly ideal liquid and solid solutions, such as methane + krypton, are called type I, according to the Bakhuis Roozeboom classification; those with a maximum melting point, an exceedingly rare type exemplified by *d*-carvoxime + *l*-carvoxime, are type II; and those with a minimum melting point (bromobenzene + iodobenzene), type III. See SOLID SOLUTION.

In most binary systems, however, extensive solid solutions are impossible because of the incompatibility of the size, shape, and crystal lattices of the two components. In the absence of solid solution formation, the addition of a solute to a liquid solvent invariably depresses the freezing point, that is the temperature at which, upon cooling, the first trace of solid solvent appears. This depression of the freezing point of a dilute solution is a convenient method (the so-called cryoscopic method) for determining the molecular weight of a solid solute. See MOLECULAR WEIGHT.

As the freezing point curve of a liquid continues to lower temperatures with higher concentrations of the second component the conventional roles of solute and solvent become reversed, and one speaks of the solubility of a solute rather than of the depression of the freezing point of the solvent; no point of demarcation exists, and the two situations are in fact only two aspects of the same phenomenon.

The second component also has a freezing-point–solubility curve marking the temperature at which the solution is in equilibrium with pure solid 2. The point of intersection of the two phase boundaries is the eutectic point which defines a eutectic temperature and a eutectic composition. Below this temperature, the system consists of two solid phases ($S_1 + S_2$ in Fig. 5*a*).

Solid-liquid phase diagrams are conveniently determined by thermal analysis of cooling curves. An initially homogeneous liquid is cooled gradually, and the temperature plotted against time. Figure 5*a* shows a simple eutectic-type phase diagram, and Fig 5*b* sketches typical cooling curves obtained for various compositions (those marked *A* to *F* in Fig. 5*a*). Curves *A* and *F* are for the

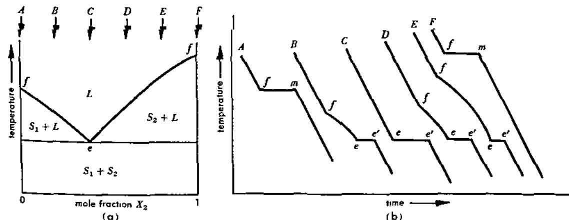


Fig. 5. (a) Schematic diagram of solid liquid equilibrium when there are no compounds of solid solutions. Points *f* are freezing points of pure substance; *e* is the eutectic point. *A*, *B*, *C*, *D*, *E*, and *F* are the compositions corresponding to the cooling curves of Fig. 5b.

(b) Schematic cooling curves for system shown in Fig. 5a, *f* = freezing point; *m* = melting point; *e* = beginning of eutectic freezing; *e'* = end of eutectic freezing. *A*, *B*, *C*, *D*, *E*, and *F* are the Fig. 5a.

pure components; each shows a single temperature arrest, a horizontal section of the curve for the melting point (between *l* and *m*), at which the temperature remains constant from the time the first bit of solid is formed until the last bit of liquid disappears. *See* THERMOANALYSIS.

Curves *B* and *E* are for solutions rich in components 1 and 2, respectively. At a temperature below the melting point of the pure substance, the first bit of solvent begins to freeze out; this is indicated by a change in slope (point *f*). The freezing section of the cooling curve (between points *f* and *e*) is not horizontal; as solvent freezes out, the liquid phase becomes richer in solute, and the freezing point is further depressed. When the composition of the liquid phase reaches the eutectic composition, the second component begins to freeze out as well as the first. No further change in liquid composition occurs, so the temperature remains constant until no liquid remains (between points *e* and *e'*). The solid which freezes out at the eutectic (the eutectic mixture) appears superficially very different from either pure solid; it is a mixture of very small crystals of each of the two components which have crystallized together. This microcrystalline two-phase mixture is in no sense a compound.

Curve *D* is a cooling curve for still another composition, and curve *C* is for the eutectic composition. Alone, curve *C* is indistinguishable from that for the freezing of a pure substance; only by combining the information from a series of cooling [illegible]

diagram can be of the eutectic type (Bakhuis Roozeboom type V) illustrated by the system silver + copper shown in Fig. 6. Indeed, since the mutual solubilities are never exactly zero, eutectic

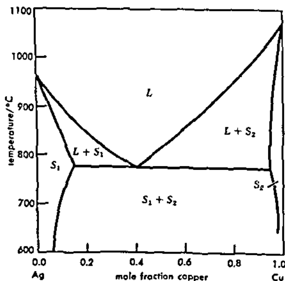


Fig. 6. Solid-liquid equilibrium in the system silver + copper (type V solid solutions).

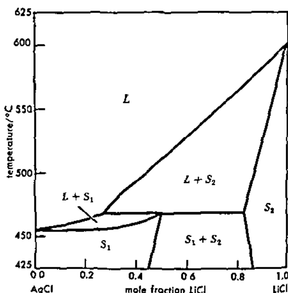


Fig. 7. Solid-liquid equilibrium in the system silver chloride + lithium chloride (type IV solid solutions).

[illegible]

a liquid phase and a solid phase of different composition, one speaks of [illegible]

(Fig. 7); similar phenomena occur in systems in which there is no appreciable solid-solution formation.

Many solid-liquid phase diagrams are complicated by the existence of intermediate crystalline phases of different crystal structure. Usually, these intermediate phases are at compositions close to simple mole ratios of the components and have the two kinds of molecules distributed in a regular arrangement; consequently, it is convenient to call these compounds, even if no specific chemical interactions can be demonstrated unequivocally. There is a rich variety of such complex phase diagrams, especially among binary systems of ionic salts ($CaCl_2$–KCl) and among binary systems of metals (alloys). Figure 8 shows the Cu + Zn system (brass), in which a whole series of crystal structures appear: α-brass has the crystal structure of pure copper (face-centered cubic) with an occasional Zn atom in the lattice; β-brass has a body-centered cubic structure with a Cu—Zn ratio of approximately 1:1; γ-brass has a very complex structure related to the formula Cu_5Zn_8; η-brass has the crystal structure of pure Zn (hexagonal, close-packed); the δ and ϵ phases have still different structures. Note the five peritectic transitions (for example, that at 600°C where the ϵ-solid melts incongruently to the δ-solid and a liquid [illegible]

Fig. 8. Temperature-composition diagram for the system copper + zinc (brass). Note the six different solid phases. The dotted line separating β from β' denotes a second-order order-disorder transition.

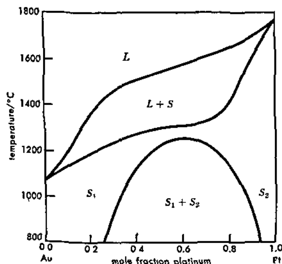


Fig. 9. Solid-liquid and solid-solid equilibrium in the system gold platinum. The critical region of the solid-solid phase boundary nearly touches the solidus curve of the solid-liquid phase boundary. If these actually touched and coalesced, one would have a type IV diagram (Fig. 7).

tures (with solid solutions of the same crystal structure) *which are analogous to the more familiar* liquid-liquid case (the system gold + platinum shown in Fig. 9). More common are the transitions between one crystalline form and another, which can occur even in pure substances (*see* TRANSITION POINT). One such occurs in the system nitrogen + carbon monoxide (Fig. 10 shows solid-solid, solid-liquid, and liquid-vapor transitions for this system).

MULTICOMPONENT SYSTEMS

As one proceeds from binary systems to systems with three or more components, the phase diagrams become more complex. Each component adds another dimension to the representation of the phase equilibria Thus, for three components, two dimensions are required to represent the phase diagrams for a single temperature and pressure; these are conveniently depicted by a triangular diagram in which each vertex represents a pure component Figure 11 shows a schematic diagram of a ternary system in which three liquid phases can coexist. Even when there are three phases, the system is still bivariant. (An example of such a system is water + succinonitrile + diethyl ether.)

A special case of a three-component system is that in which there are two immiscible solvents and a third component, soluble in both, distributed between the two phases. The ratio of the concentrations of the solute in the two solvents is the distribution coefficient; in dilute solutions, this is independent of the concentration, but at higher *concentrations, nonideal behavior of the solute can* produce systematic variations of the distribution coefficient which ends in the most concentrated solutions with the ratio of the solubilities in the rated solutions.

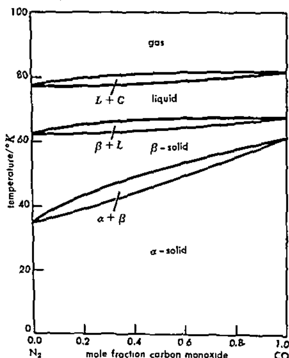


Fig. 10. Phase equilibria in the system nitrogen + carbon monoxide, showing solid-solid, solid-liquid, and liquid-vapor equilibrium.

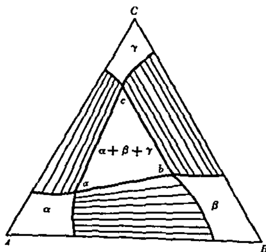


Fig. 11. Schematic diagram of a three-component system at a fixed temperature and pressure. Points A, B, and C represent the pure liquids. The composition corresponding to a point in the diagram is determined by the positions along a line from each vertex to the opposite side; thus point b is 20% A, 50% B, 30% C. Regions α, β, and γ correspond to single phases rich in A, B, and C, respectively; $\alpha + \beta + \gamma$ is a three-phase region; the three saturated solutions have the compositions given by the points a, b, and c. Three two-phase regions, $\alpha + \beta$, $\alpha + \gamma$, and $\beta + \gamma$, are indicated by drawing in the tie lines.

Distribution effects are important in separating similar materials. A small difference in distribution coefficients is amplified by multistage equilibria, such as those used in countercurrent extraction and partition chromatography.

The examples used to illustrate the various types of phase equilibria are not supposed to suggest that these types are restricted to the particular kind of chemical substances shown.

elements, inorganic salts, and organic nonelectrolytes. *See* ALLOY STRUCTURES; COUNTERCURRENT MASS-TRANSFER OPERATION; CRYSTAL STRUCTURE; EQUILIBRIUM, CHEMICAL; EXTRACTION; FUSED-SALT PHASE EQUILIBRIA; INTERFACE OF PHASES; SILICATE PHASE EQUILIBRIUM; SULFIDE PHASE EQUILIBRIA; THERMODYNAMICS (CHEMICAL). [R.L.S.]

Bibliography: A. Findlay, (A. N. Campbell and N. O. Smith, eds.), *The Phase Rule and Its Applications*, 9th ed., 1951; J. H. Hildebrand and R. L. Scott, *Solubility of Nonelectrolytes*, 3d ed., 1950; J. S. Marsh, *Principles of Phase Diagrams*, 1935; G. Masing, (B. A. Rogers, trans.), *Ternary Systems*, 1944, J. E. Ricci, *The Phase Rule and Heterogeneous Equilibrium*, 1951; H. W. Bakhuis Roozeboom (ed.), *Die heterogene Gleichgewichte vom Standpunkte der Phasenlehre*, 3 vols., 1901–1918; F. E. W. Wetmore and D. J. LeRoy, *Principles of Phase Equilibria*, 1951.

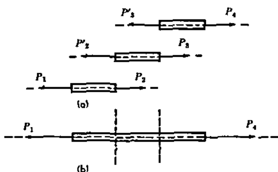
Equilibrium of forces

In a mechanical system the condition under which no acceleration takes place. Newtonian mechanics today is based upon two definitions which modify, but are essentially equivalent to, Newton's three fundamental laws. These definitions postulate the action of forces on particles. A particle is defined as a conceptual volume element that has mass and is sufficiently small to have point location. A body is defined as a system of particles. To develop the mechanics of a body, these definitions are applied to each of its particles and their influences summed. *See* PARTICLE.

Newtonian particle laws. The law of motion is that in a Newtonian frame of reference a particle of mass m acted on by resultant force $\mathbf{F}$ has acceleration $\mathbf{a}$ in accordance with the equation $\mathbf{F} = km\mathbf{a}$. Therein, k is a positive constant whose value depends upon the units in which $\mathbf{F}$, m, and $\mathbf{a}$ are measured.

The action-reaction law states that when one particle exerts force on another, the other particle exerts on the one a collinear force equal in magnitude but oppositely directed.

Equilibrium of a particle. A particle at rest or in uniform (unaccelerated) motion in a Newtonian reference frame is in equilibrium. With few exceptions a frame of reference fixed with respect to earth is considered to be Newtonian and the equation $\mathbf{F} = km\mathbf{a}$ applied. Accordingly, when a particle is in equilibrium on earth, $\mathbf{a} = 0$ and thus



Equilibrium forces. (a) Action and reaction forces on particles of a body. (b) Balanced external forces on a body.

F = 0. Also, when **F** = 0, **a** = 0 because k and m are not zero. These considerations lead to a theorem and its corollary: The resultant of forces exerted on a particle in equilibrium is zero; and if the resultant of forces exerted on a particle is zero, the particle is in equilibrium.

Body equilibrium. A body acted upon by force is in equilibrium when its constituent particles are in equilibrium. The forces exerted on its particles (and therefore on the body) are either internal or external to the body. An internal force is one exerted by one particle on another in the same body.

An external force is one exerted on a particle or the body by a particle not of the body.

Being particle actions, internal forces obey the action-reaction law. They occur in pairs whose individual and combined resultants are zero. Hence a further theorem is that the resultant of all forces internal to a body is zero.

The resultant of forces on each particle and therefore on all particles of a body in equilibrium is zero. Because the forces are either internal or external and the resultant of internal forces is zero, it follows as a theorem that the resultant of all external forces acting upon a body in equilibrium is zero.

Equations of equilibrium. Forces whose resultant is zero have zero vectors of total force and total moment (*see* RESULTANT OF FORCES). Consequently, both the algebraic sum of their components in any direction and their sum of moments about any line are also zero. Specifically, relative to an $OXYZ$ reference frame, the components and moments of forces externally applied to a body in equilibrium are related through the six equations

$$\Sigma X = 0 \quad \Sigma M_x = 0$$
$$\Sigma Y = 0 \quad \Sigma M_y = 0$$
$$\Sigma Z = 0 \quad \Sigma M_z = 0$$

Dependence of equations. Forces can be collinear, coplanar, parallel, concurrent, and so forth. According to the class of a particular external force system and its orientation in $OXYZ$, certain of these equilibrium equations may be either trivial (containing no terms on their left sides), or dependent (derivable by arithmetic operations on one or more previously written equations of the set). The number of equations that are independent (contain different information) is the number necessary to define the resultant of the particular class force system which acts.

Force interaction of bodies. Under certain conditions, particles of one body exert pertinent forces on particles of another. Such forces, being particle actions, obey the action-reaction law. Thus, the entire force system exerted by one body on another coexists with an equal but opposite system exerted by the other on the one. The resultants of the coexistent systems are also equal and opposite. These conditions lead to the theorem that the mutual force actions of bodies are equal and opposite.

Contact and body forces. External forces are either contact or body forces according to their manner of application.

Contact forces are those exerted between the contacting particles of bodies which touch physically. The resultant of such forces is a force, couple, or force and couple, depending partly upon the nature of the contacting surfaces. For example, rigid bodies whose contacting surfaces do not penetrate, interlock, or adhere, exert resultant compressive (push) forces on one another which act at the point or within the area of contact. If, additionally, either surface is considered to be smooth (incapable of resisting or applying force tangent to the surface) the force is normal to the surfaces at the contact.

Body forces are those exerted upon certain or all particles of a body by matter which need not contact the body. Gravitational, magnetic, and electrostatic forces are examples. A body's weight is the force resultant of essentially parallel downward forces resulting from the gravitational pull of the earth on each particle of the body. The weight force acts downward through the geometric center of a homogeneous body. *See* CENTER OF GRAVITY.

Problem of equilibrium. To determine the forces which act upon stated matter at rest on earth is the usual problem of equilibrium. The problem is statically determinate if solution is possible by employing the equations of equilibrium only. These equations relate all external forces; hence accurate solution requires that external forces be recognized and described according to their agencies of application. To summarize this information the free-body diagram is conveniently employed. A free-body diagram is the sketch of a body of interest with all external forces acting thereon symbolized by arrows drawn on the sketch. *See* DYNAMICS; FORCE; KINETICS (CLASSICAL MECHANICS); STATICS.

[N.S.F.]

Equinox

The date of the year and the point on the celestial sphere at which the sun's rays at noon are 90° above the horizon at the Equator, or at an angle of

90° with the earth's axis, and neither North nor South Pole is inclined toward the sun. This phenomenon occurs on two days of the year, approximately March 21 and September 23. In the Northern Hemisphere the event in March is referred to as the vernal equinox, and that in September as the autumnal equinox. As conditions are the same on the earth at both the vernal and autumnal equinoxes, no distinction need be made here between the two. *See* ECLIPTIC.

With the sun's rays at 90° to the earth's axis the rays will illuminate half the globe, from pole to pole, thus bisecting all parallels. Therefore the sun will appear at all places on the earth and will give 12 hours of sunlight and 12 hours of darkness, not considering twilight. In other words, at all places, the sun at the equinox will appear in the east at 6 A.M., local sun time, and set at 6 P.M., in the west; thus the term equinox, means equal night.

The situation at the two poles is slightly different; there on this day the sun will rise on the eastern horizon and will make one circuit of the horizon in 24 hours. On the day after the vernal equinox the sun will rise higher and remain above the horizon at the North Pole, but will drop below the horizon at the South Pole. The reverse is true following the autumnal equinox, when the sun remains above the horizon at the South Pole. The vernal equinox is considered the first day of spring in the Northern Hemisphere, and the autumnal equinox the first day of fall.

At the time of the equinox the altitude of the noon sun at any place is the same as the colatitude (90° minus the latitude) of the place. Thus at the Equator it is 90°, at the poles 0°, and at 45° north or south, it is 45°. *See* ASTRONOMICAL COORDINATE SYSTEMS; GEOGRAPHY, MATHEMATICAL; PRECESSION OF EQUINOXES. [V.H E.]

Equisetales

An order of the class Equisetineae of the subphylum Sphenopsida, phylum Tracheophyta. Equisetales, commonly known as horsetails, is represented by a single living genus, *Equisetum*, with about 25 species found both in moist and dry habitats. These plants grow throughout the world except in Australia and New Zealand. The plants range from herbaceous to shrubby and rarely exceed 3 or 4 ft in height, although some tropical species grow much taller.

Structure. The plant body is commonly composed of perennial underground stems (rhizomes) with various types of aerial stems (Fig. 1). Some of these are perennial, others annual; some are unbranched and reproductive, others much branched and vegetative (*see* ANNUAL PLANTS; PERENNIAL PLANTS). The bushy structure of the latter is suggestive of the common name, horsetail.

The leaves are reduced to scales which appear fused in whorls at the nodes. In the young leaf, the central region contains stomates and chloroplasts which function in photosynthesis for a short time.

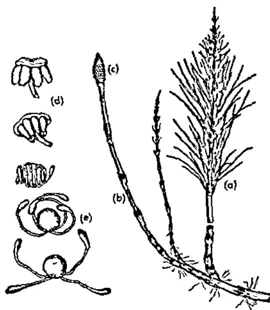


Fig. 1. *Equisetum arvense.* (a) Sterile shoot. (b) Fertile shoot growing from an underground rootstock. (c) Cone. (d) Two views of shield-shaped sporangiophores. (e) Spores, greatly enlarged. As the spore dries, the elaters expand. (*From E. W. Sinnott and K. S. Wilson, Botany: Principles and Problems, 5th ed., McGraw-Hill, 1955*)

The burden of photosynthesis is carried on by cells in the cortex of the stem. There are no leaf gaps in the stele. *See* LEAF (BOTANY).

The stem is composed of conspicuous nodes with long, strongly ribbed internodes. *See* STEM (BOTANY). The cells of the stem originate in a single, pyramidal apical cell giving rise eventually to a ring of vascular bundles with numerous openings or canals present except at the node (*see* MERISTEM, APICAL; VASCULAR BUNDLES). These bundles are arranged into an endarch siphonostele (*see* STELE). The walls of sclerenchyma fibers in the cortex and of the epidermal cells are more or less impregnated with silica and aid in giving mechanical support to the stem (*see* CORTEX, PLANT; EPIDERMIS, PLANT; SCLERENCHYMA). Early settlers, noting this, made use of the vegetative plants for scouring pots, pans, and floors, and as a result gave them the erroneous common name of "scouring rushes."

Except the first-formed primary root, all roots are adventitious. *See* ROOT (BOTANY). Each adventitious root is derived from a single apical cell and each develops an exarch protostele. There is no true cambium, but an intercalary meristem (region of cell division) has been observed. The strobili (cones) are solitary and terminal, either on the main stem or on lateral branches. They are composed of a central axis with a whorled series of compact lateral branches known as sporangiophores. Sporophylls (spore-bearing leaves) are entirely suppressed. Because of the pressure resulting from their development, each sporangiosphore

appears compressed and hexagonal in surface view. At maturity the several sporangia ripen on the lower or abaxial portion of the sporangiophore. The spores produced are alike (homosporous). Each possesses a complex wall structure having four long, slender appendages known as elaters. These elaters are hygroscopic (coil and uncoil in response to changes in humidity) and thus aid in dissemination of spores. Spores which reach a suitable substrate germinate immediately, producing a flat plate or mound of green tissue with numerous filamentous branches called the gametophyte (Fig. 2) which usually gives rise to the sex organs, antheridia (male) and archegonia (female). The motile sperms produced in the antheridia are large, spirally coiled bodies with numerous apical flagella. The archegonia are formed in the cushionlike portion of the gametophyte, with the chambers containing the eggs well buried in the thallus. Several eggs on the same gametophyte may be fertilized, following which each may develop an embryonic sporophyte (Fig. 3) without a suspensor (a chain of cells which serves to put the embryo in a favorable position in relation to its food supply). With the formation of the primary root and stem, the young sporophyte produces chlorophyll and becomes independent (self-sustaining). *See* EQUISETINEAE; SPHENOPSIDA; TRACHEOPHYTA. [P.A.V.]

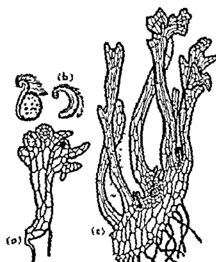


Fig. 2. Gametophyte of Equisetum. (a) Small gametophyte, showing antheridia only (after Hofmeister). (b) Male gametes (after Schacht). (c) Larger gametophyte, showing the long, branching lobes with archegonia at their bases (after Hofmeister). (From E. W. Sinnott and K. S. Wilson, Botany: Principles and Problems, 5th ed., McGraw-Hill, 1955)

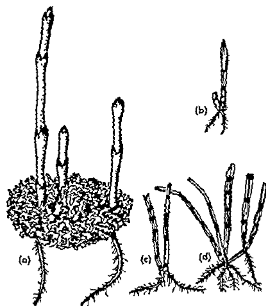


Fig. 3. *(a) Equisetum sp. Gametophyte with young sporophytes. (b, c, d) Equisetum hiemale L., young sporophytes (after E. C. Jeffrey, 1899). (From G. M. Smith, Cryptogamic Botany, vol. 2, 2d ed., McGraw-Hill, 1955)*

Bibliography: H. C. Bold, *Morphology of Plants*, 1957; H. J. Fuller and O. Tippo, *College Botany*, rev. ed., 1954; A. W. Haupt, *Plant Morphology*, 1953; G. M. Smith, *Cryptogamic Botany*, vol. 2, 2d ed., 1955.

Equisetineae

The only class of the plant subphylum Sphenopsida. This class is usually divided into three orders: (1) Hyeniales, known only by the fossil genera *Hyenia* and *Calamophyton*; (2) Sphenophyllales, known only by the one fossil genus *Sphenophyllum*, and (3) Equisetales, known by the fossil genus *Calamites* and the living genus *Equisetum*, commonly called horsetails. This line of evolution seems to parallel that of the subphylum Lycopsida in that it had its beginnings in the Devonian, its peak in the Carboniferous, and exists today only as a remnant group of living plants. It is of particular interest because of its evolutionary history and because, with the Lycopsida, its fossil members contributed greatly to the Carboniferous coal seams. *See* EQUISETALES; LYCOPSIDA; SPHENOPSIDA; TRACHEOPHYTA; *see also* CALAMITALES; HYENIALES; PALEOBOTANY; SPHENOPHYLLALES. [P.A.V.]

Bibliography: *See* EQUISETALES.

Equivalent circuit

A representation of an actual electric circuit or device by a simplified circuit whose behavior is identical to that of the actual circuit or device over a stated range of operating conditions. These conditions depend on the elements contained in the actual circuit and may include such variables as frequency, temperature, and pressure, in addition to voltage and current. Equivalent circuits are often used by engineers to simplify circuit analysis . . . they show the relation between the variabl

clearly than the actual circuit. *See* NETWORK THEORY, ELECTRICAL.

Two types of equivalent circuits are used. One type is a simplification of an actual linear device, such as a transformer. In general, there are no restrictions on the magnitudes of applied voltages and currents in this type of equivalent circuit; therefore the equivalence holds over a wide range of operating conditions. *See* TRANSFORMER.

The second type of equivalent circuit is used to represent nonlinear devices, most commonly vacuum tubes and transistors. The actual circuit of the nonlinear device can be replaced by a fictitious generator and a simple network representing the operation of the actual device. Both the generator and network quantities depend on the parameters of the device. These parameters vary; therefore calculations made are necessarily limited to a portion of the characteristics over which the values of the *parameters are reasonably constant*. This type of circuit can, therefore, be used only for analysis of small signals. For large-signal analysis of nonlinear devices, it is necessary to use graphical methods on the characteristic curve of the device.

[R.L.R.]

Equivalent weight

The number of parts by weight of an element or compound which will combine with or replace, directly or indirectly, 1.008 parts by weight of hydrogen, 8.00 parts of oxygen, or the equivalent weight of any other element or compound. The term equivalent weight comes from the law of equivalent proportions which states that the weights of two elements A and B which combine separately with identical weights of another element C are either the weights in which A and B combine together, or are related to them in the ratio of small whole numbers. A standard weight of 8.000 parts is chosen for oxygen. For all elements the atomic weight is equal to the equivalent weight times a small whole number, called the valence of the element. *See* CHEMICAL COMPOUNDS; ELEMENTS (CHEMICAL); VALENCE.

An element can have more than one valence and therefore more than one equivalent weight. The use of the terms is illustrated below.

1. Ammonia, NH_3, contains 1 atom of nitrogen combined with 3 atoms of hydrogen. Since the equivalent weight of hydrogen is equal to its atomic weight, the equivalent weight of nitrogen is ⅓ its atomic weight and its valence is 3.

2. Magnesium oxide, MgO, contains 1 atom of magnesium combined with 1 atom of oxygen. Since the equivalent weight of oxygen is ½ its atomic weight, the equivalent weight of magnesium is also ½ its atomic weight and its valence is 2.

3. Phosphorus forms two chlorides, phosphorus trichloride, PCl_3, and phosphorus pentachloride, PCl_5. Since the equivalent weight of chlorine is equal to its atomic weight, in the trichloride the equivalent weight of phosphorus is ⅓ its atomic weight and its valence is 3, and in the pentachloride the equivalent weight is ⅕ its atomic weight and its valence is 5.

The equivalent weight of a compound depends on the reaction in which it takes part. Thus:

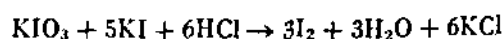
1. In the reaction between potassium iodate, KIO_3, and silver nitrate, $AgNO_3$, one molecule of silver iodate, $AgIO_3$, is precipitated for every molecule of silver nitrate.



$$KIO_3 + AgNO_3 \rightarrow AgIO_3(ppt) + KNO_3$$

Since the equivalent weight of silver is its atomic weight, the equivalent weight of potassium iodate, in this reaction, is its molecular weight.

2. When potassium iodate, KIO_3, is reduced to iodine, I_2, by potassium iodide, KI, three molecules of iodine are produced per molecule of potassium iodate.



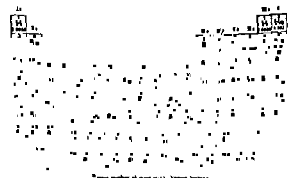
$$KIO_3 + 5KI + 6HCl \rightarrow 3I_2 + 3H_2O + 6KCl$$

Since the equivalent weight of iodine is ½ its molecular weight, the equivalent weight of potassium iodate, in this reaction, is ⅙ its molecular weight. *See* ATOMIC WEIGHT; ELECTROCHEMICAL EQUIVALENT; OXIDATION-REDUCTION.

[T.C.W.]

Erbium

Element number 68, erbium, Er, is a metallic element belonging to the rare-earth group. Its atomic weight is 167.27, and the naturally occurring element is made up of the stable isotopes Er^{162} 0.136%, Er^{164} 1.56%, Er^{166} 33.41%, Er^{167} 22.94%, Er^{168} 27.07%, Er^{170} 14.88%. It was discovered in 1843 by C. G. Mosander, who originally named this oxide terbia. The name was changed to erbium in 1860. Erbium forms a rose-pink oxide, Er_2O_3, which dissolves in mineral acids to give rose-colored solutions. The salts are paramagnetic and the ions are trivalent. For properties of the metal, *see* RARE-EARTH ELEMENTS. At low temperatures, the metal is antiferromagnetic and at still lower temperatures becomes strongly ferromagnetic. The Néel point is 78°K and the Curie point about 20°K.



[F.H SP.]

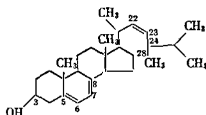
Erg

A unit of energy or work in the centimeter-gram-second system of units. It is equal to the work done by a force of magnitude 1 dyne when the point at

should never be used as a synonym for "dyne-centimeter of torque." *See* WORK; *see also* UNITS, SYSTEMS OF. [D.F.R.]

Ergosterol

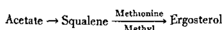
A sterol which is the most common of the mycosterols, the sterols of yeast and fungi (*see* STEROL). The name is derived from the fact that it was first discovered in ergot bodies, on rye and other cereal plants diseased with ergot fungi. Ergosterol is a white, crystalline substance, insoluble in water and soluble in organic solvents. The structural formula is



Like cholesterol, it contains a hydroxyl group of the β configuration at carbon atom 3, but has three positions of unsaturation, at carbon atoms 5–6, 7–8, and 22–23. It differs from cholesterol also in containing a methyl group at carbon atom 24, instead of a hydrogen atom (*see* CHOLESTEROL).

Ergosterol is obtained in large quantities, by cultivation of certain strains of yeast (*see* YEAST, INDUSTRIAL). Irradiation with ultraviolet light leads to the formation of lumisterol, tachysterol, calciferol, and other compounds, by varied arrangements of the molecule. Calciferol is Vitamin D_2 This process, which under controlled conditions gives a yield of the vitamin of more than 50%, is used commercially. *See* VITAMIN D.

The biosynthesis of ergosterol in yeast proceeds by the same or a similar pathway to that of cholesterol in animal tissues:



$$\text{Acetate} \rightarrow \text{Squalene} \xrightarrow[\text{Methyl}]{\text{Methionine}} \text{Ergosterol}$$

Both acetate and squalene are converted to ergosterol by enzymes in yeast. The methyl group of the amino acid, methionine, serves as a source of the methyl group, on carbon 28, which becomes attached to carbon atom 24 of ergosterol *See* STEROID. [I.Z.]

Ericales

An order of the plant subclass Dicotyledoneae including 6 families with 114 genera and about 2400 species. This is the primitive order of the Sympetalae; some of the lower forms are polypetalous, and in others the stamens are free from the corolla. The anthers open by terminal pores in the tubular prolongations of the pollen sacs. The plants, mostly shrubs and trees, commonly called heaths, are characteristic of northern latitudes. The Lennoa family (Lennoaceae) is a small group of parasitic herbs occurring in the southwestern United States and in Mexico. The heath family (*Ericaceae*) is the dominant family, economically valuable in landscaping and for edible fruits. This family includes the rhododendrons and the blueberries. *See* BLUEBERRY; *see also* DICOTYLEDONEAE; EMBRYOPHYTA; PLANT KINGDOM. [P.D.S.]

Eros

An asteroid discovered by G. Witt at Berlin in August, 1898. It has played an important role in astronomy because of its occasional proximity to Earth and the application of this circumstance to the determination of the solar parallax, which provides the value of the so-called astronomical unit (AU) of distance, that is, the mean distance of the Earth to the Sun. *See* PARALLAX (ASTRONOMY).

The semimajor axis of its orbit or mean distance to the Sun is $a = 136 \times 10^6$ miles; the eccentricity $e = 0.223$ causes the actual distance to vary from 106×10^6 miles at perihelion to 165×10^6 miles at aphelion; the inclination of the orbit to the plane of the ecliptic is $i = 10.8°$, and the sidereal period of revolution $P = 643.2$ days.

At its closest approach to Earth the distance of Eros is only 14×10^6 miles, about half the minimum distance of Venus, and its parallax is then almost 60″. Favorable oppositions of Eros in 1901 and 1931 were extensively observed for the determination of the solar parallax; the next one will occur in 1975.

Eros is probably an irregular, elongated object, perhaps 20 miles long and 5–10 miles wide, rotating around its short axis in 5h 16m. *See* ASTEROID; PLANET. [G.D.V.]

Bibliography: H. N. Russell, R. S. Dugan, and J. Q Stewart, *Astronomy*, rev. ed., vol. 1, 1945.

Erosion

The loosening and transporting of rock debris at the earth's surface, aptly described as the "wearing away" of the land. Agents of erosion include surface, ground, and ocean water; ice (especially glaciers), wind; gravity; and organisms. *See* WEATHERING PROCESSES.

Erosion removes, on the average, 1 ft (vertically) of rock material in the order of thousands of years From the time it begins until the lowest possible level has been reached, erosion progresses through stages empirically described as youth, maturity, and old age, each characterized by a distinctive group of landforms (*see* GEOMORPHOLOGY).

Erosion is of great concern to man because it removes the fertile topsoil, changes watercourses and landforms, and causes damage to valuable man-made structures. Erosion has removed valuable ore deposits and rendered some land uninhabitable, but has also stripped off worthless overburden and smoothed wide areas, making some mineral deposits available, and making land suitable for agricultural purposes. Quickening of the pace of erosion (accelerated erosion), brought about by

escalator transports a continuous stream of passengers from floor to floor; its constant motion is an invitation to passengers to use its facilities. Usually speed is 90 ft/min; slope is standardized at 30°. A 32-in wide escalator can carry 85 passengers per minute; a 48-in. installation has a capacity of 135 passengers per minute. Steps ride on resilient rollers running on tracks; endless roller chains propel the steps, the chains being driven by sprockets and a worm drive. A 10- or 15-hp electric motor at the top of the escalator provides the motive force. Direction of travel can be reversed in accordance with traffic flow; usually both up and down flights are installed. [F.H.R.]

Escape velocity

Minimum speed away from a parent body that a particle must acquire to escape permanently from the gravitational attraction of the parent. Escape velocity is also termed parabolic velocity (*see* ORBITAL MOTION).

Earth retains an atmosphere because the escape velocity is considerably higher than the mean velocity of the gas molecules in its atmosphere. For a space ship to escape from Earth and travel to another planet or orbit about the Sun, it must reach escape velocity. This velocity v can be calculated by equating the kinetic energy of the moving body (gas molecule, space ship, or other matter) of mass m to the work necessary to overcome the gravity g_0 at the surface of the parent whose radius is r_0 in rising to a height h above the surface of the parent; thus

$$\frac{mv^2}{2} = m \int_0^h g_0 \left(\frac{r_0}{r_0 + h}\right)^2 dh = \frac{mg_0r_0h}{r_0 + h}$$

from which

$$v = \sqrt{2g_0r_0}\sqrt{\frac{h}{r_0 + h}}$$

where for Earth $r_0 = 6.38 \times 10^6$ meters and $g_0 = 9.8$ m/sec² and for escape $h = \infty$ so that the term under the second radical is unity, thus $v_{escape} \approx 11.2 \times 10^3$ m/sec. For comparison, the mean velocity of a gas with molecules of mass m at absolute temperature T is $v = \sqrt{3kT/m}$ where k is Boltzmann's [illegible]

[illegible] for other bodies in the solar system in meters per second are Mercury, 3.8×10^3; Venus, 10.4×10^3; Mars, 5.1×10^3; Jupiter, 61.0×10^3; Saturn, 36.7×10^3; Uranus, 22×10^3; and Neptune, 24×10^3. *See* CELESTIAL MECHANICS; SATELLITE, ARTIFICIAL. [R.L.D.]

Escapement

A mechanism comprising a toothed wheel that engages alternate pallets attached to a vibrating (oscillating) member. The escapement is found principally in timekeepers but may be employed wherever oscillating motion is required.

In a clock or watch, the escapement intervenes between the energy source (spring or elevated weight) and the regulating device (pendulum or balance wheel). It is acted upon by each. The escape wheel is mounted on the same shaft as the last wheel of the gear train, and impulses are delivered from the escape wheel to operate the regulating device. The regulating device, which has a natural period of vibration, determines the rate at which it will receive these impulses, and thus regulates the rate of going of the timepiece.

Anchor recoil escapement (Fig. 1) is used with a pendulum and takes its name from the shape of its oscillating member and its action. This type of escapement appeared late in the seventeenth century and has survived, with modifications, to the present. To simplify the explanation, the pendulum is shown attached directly to the anchor, so that the pendulum and anchor swing as one.

In the position shown, pallet B is receiving an impulse from the escape wheel which swings the pendulum to the left. When pallet B has cleared the tooth, allowing the wheel to "escape," pallet C will be in position to arrest the wheel. Recoil, or momentary reversal of the escape wheel, occurs just after it is arrested because the pendulum has not quite completed its swing. The wheel tooth in contact with pallet C will then give the oscillating parts an impulse in the opposite direction.

Dead beat, an escapement without recoil, has arresting faces of the pallets described by circular arcs whose centers are at the pivot point of the anchor. Escape wheel teeth are contoured to give impulses to these pallet faces.

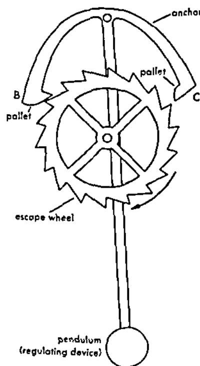


Fig. 1. Anchor recoil escapement.



Fig 2. Detached lever escapement. (F. J. Britten, *Britten's Old Clocks and Watches and Their Makers*, 7th ed., Dutton, 1956)

Modern watches generally employ a detached lever escapement (Fig. 2), which has banking pins *B* to limit the oscillation of the anchor and its lever. An escapement is termed detached when the regulating device, in this case the balance wheel, is given an impulse during only a small part of its operating cycle. When the fork reaches the end of its swing it is lightly locked by a wheel tooth and remains stationary until the returning impulse pin *E* causes sufficient recoil of the escape wheel to release the pallet. The chronometer escapement is a detached escapement that furnishes an impulse in only one direction of swing of the balance wheel.

Escapements are also adapted for use in automatic factory equipment to regulate the flow of parts from magazine or feeder track. Such feeding and spacing devices are of various forms, many being adaptations of the ratchet and pawl (*see* RATCHET). [E.S.F.]

Bibliography: K. R. Treer, Escapements for automatic equipment, *Automation*, pp. 80–86, Feb. 1957.

Escarpment

A cliff or steep slope of some extent, generally the margin of a plateau or the steep face of an asymmetrical ridge. Escarpments, or scarps, may be produced by faulting, erosion, or sapping of less resistant underlying strata to create cliffy rock faces in massive layers above, as on the walls of the Grand Canyon of the Colorado River. Spectacular waterfalls may plunge over scarps, as along the glacially-steepened sides of Yosemite Valley where jointing of granitic rock determines the locations of cliffs. The north face of the San Jacinto Range in southern California is the steepest and highest fault scarp in the United States. *See* FAULT AND FAULT STRUCTURES.

More commonly scarps are lower and result from differential rates of weathering and erosion of contrasted rock types. The denudation of inclined strata may produce cuestas (cuesta escarpments) or hogbacks with dip slopes just as steep as the escarpment. Both have backslopes which approximate the dip of their sedimentary layers and steeper scarps which truncate the bedding. Niagara Falls is supported by a cuesta of erosionally resistant limestone which crops out for several hundred miles along an arc from New York through Ontario, and which eventually swings southward through Wisconsin into Illinois. Low scarps may occur at the boundary between older, more resistant bedrock and younger, less resistant strata. A hidden scarp east of Ozark Plateau, which lies buried beneath alluvial fill, once attained heights of as much as 80 ft. Many coastal plains exhibit low cuestas with scarps directed inland for the reason that their sedimentary strata dip seaward. Some are called wolds in the South because they were originally densely wooded, as was the English weald, or the Schwarzwald of Germany. *See* COASTAL PLAIN; FALL LINE. [R.J.RU.]

Escherichia

A genus of bacteria of the family Enterobacteriaceae (*see* ENTEROBACTERIACEAE). The type species, *Escherichia coli* (*Bacterium coli*), a normal commensal of the intestinal tract of man and animals, also causes infections of the urinary tract, localized infection of the appendix, gallbladder, and rectum.

The members of the genus are gram-negative, rod-shaped bacilli usually motile by peritrichous flagella and lack proteolytic enzymes. The methyl red test is positive. The Voges-Proskauer test is negative (*see* IMViC TEST). The lactose fermenting colon bacillus is often distinguished from the non-lactose-fermenting paracolon bacillus. Coliform bacilli is a widely used loose term, usually employed to include members of the genera *Escherichia* and *Aerobacter* (*see* AEROBACTER).

A considerable number of serological types can be distinguished by antigenic analysis (*see* IMMUNOLOGY). Three antigenic components are employed for this classification: somatic (O) antigens, capsular or envelope (K) antigens, and flagellar (H) antigens. For the principles of this procedure *see* SALMONELLA.

E. coli is part of the normal flora of the intestinal tract and its presence in drinking water is considered an indication of fecal contamination, either animal or human (*see* WATER MICROBIOLOGY). The *Escherichiae* are common agents of infection of the urinary tract, and cause localized purulent infections in such areas as the gall bladder, the appendix, and the rectum, which may cause septicemic infections, particularly in debilitated persons (*see* SEPTICEMIA).

Certain serological types of *E. coli* are found connected with epidemics of infant diarrhea. The infant disease as well as occasional adult diarrhea

may also be caused by related bacteria of uncertain taxonomic position, intermediate between *Escherichia* and *Shigella*, *Bacillus alkalescens* and *B. dispar*. According to presently prevalent opinion, they are considered as *Escherichiae*. *See* BACTERIOLOGY, MEDICAL. [A.J.W.]

Bibliography: P. R. Edwards and W. H. Ewing, *Identification of Enterobacteriaceae*, 1955.

Esker

A ridge of poorly sorted glacial gravel and sand, varying in height from a few feet to several tens of feet and in length from a small fraction of a mile to scores of miles. The trend is roughly parallel to ice motion and the terminus may be higher than the source. Eskers are most abundant in moraine-covered terranes with slight relief but may cross divides with moderate elevations. Their form is complicated by multiple ridges, gaps, adjacent depressions, and mounds of similar material. Eskers should be distinguished from crevasse fillings which are outwash between ice blocks. True eskers were formed by streams of glacial meltwater confined between ice walls or flowing locally in tunnels. Their courses indicate that the glacier was stagnant when the eskers formed. *See* GLACIATED TERRANE. [F.T.T.]



An esker ridge in Wisconsin. This transverse contour view looks along the ridge summit toward the center background. (*Photograph by W. C. Alden, U.S. Geological Survey*)

Esophagus

A section of the alimentary canal that is interposed between the pharynx and the stomach. Because of divergent specializations in the various vertebrates the esophagus cannot be described in general terms and is not always distinguishable.

In man it is a tube running the full length of the neck and the thorax, held in its position ventral to the vertebral centra by a tunica adventitia of loose connective tissue. It has an inner lining of folded mucous membrane with an exceptionally thick lamina propria, a submucosa of elastic and collagenous connective tissue, and two layers of muscle. The musculature is striated in the anterior third of its length, unstriated in the posterior third, and variably intermixed in the middle. It is supplied with autonomic nerve fibers.

Although normally collapsed, the human esophagus is capable of considerable distention during the rapid passage of swallowed material, under which condition the folds of mucous membrane and lamina propria are temporarily smoothed out. Numerous microscopic esophageal glands open into the lumen, extending their compound tubules out into the submucosa.

In man the transition from the esophagus to the stomach occurs quite abruptly at the diaphragm. In many vertebrates the distinction between the two is not so clear. The pharynx narrows posteriorly like a funnel and the foregut may thereupon enlarge, but much of what appears to be stomach may have an esophageal character histologically. In a number of fish species there is no stomach enlargement at all. There is a metabranchial foregut anterior to the entrance of ducts from the liver and pancreas, but application to it of the names stomach and esophagus is of questionable value. At the other extreme, in some sharks and bony fishes there is a constricted esophagus with a sparsely glandular or cornified lining. The esophagus of some turtles is studded with harsh horn-tipped papillae. In some of the carnivorous birds the esophagus may be temporarily dilated for the storage of large masses of swallowed food before they can be digested. Many graminivorous birds are provided with a permanent ventral esophageal pouch, the crop, for that purpose. Pigeons of both sexes feed their newly hatched young with a thin paste consisting of cells proliferated and sloughed from special areas of the crop under the stimulus of the hormone prolactin. This is pigeon's milk. [W.W.B.]

Esophagus, disorders of

Any pathology of the esophagus, the canal extending from the pharynx to the stomach, including disorders of congenital, inflammatory, vascular, neoplastic, and mechanical origin.

Congenital defects are infrequent, but stenosis, a narrowing, and atresia, a failure of development, are seen with some regularity on large pediatric services in hospitals. Many cases are amenable to surgery but prompt diagnostic evaluation is necessary, particularly in the newborn who cannot feed properly.

Esophagitis, or inflammation, most commonly follows the accidental swallowing of corrosive agents by children. Other causes are infections, persistent vomiting, and reflux of stomach acid into the esophagus. Usually a superficial ulceration is produced and is of little significance unless erosion of blood vessels and hemorrhage follow. Corrosives may produce very severe or fatal damage, depending upon the agent, the length of exposure, and the complications which follow.

Vascular lesions are most often produced by certain forms of cirrhosis in which varicose veins around the esophagus occur as the result of portal hypertension. These varices may rupture, causing hemorrhage, inflammation, or ulceration, of acute or chronic forms. More than 10% of patients with cirrhosis will bleed from esophageal varices at some time during their illness, although the bleeding may not be grossly evident.

Carcinoma of the esophagus is the most important of the neoplastic processes, and it accounts for more than 5% of all malignancies other than cutaneous neoplasms in men. It occurs most often in persons over 50 years of age and in any case of dysphagia, or difficulty in swallowing, carcinoma must be excluded before another diagnosis is made. Obstruction, production of esophageal-tracheal fistulae, and hemorrhages are the most frequent complications.

Benign tumors, such as fibromas and lipomas, are not uncommon but they rarely produce symptoms and are generally of little clinical significance, because of their small size.

Mechanical lesions include diverticula which may be of two main types. Traction diverticula are outpouchings of the esophageal wall caused by adhesions following previous inflammatory disease of the adjacent lymph nodes or mediastinal tissue, usually following tuberculosis. Pulsion diverticula are small to large pouches which extrude through surrounding tissues, such as the pharyngeal muscles or the diaphragm. These are thought to result from developmental defects in the walls or overlying tissues through which a portion of the esophagus herniates.

Small diverticula may not cause symptoms, or if present, symptoms may be mild or transitory in nature, depending upon other conditions such as associated disease.

The two major symptoms of esophageal disorders are dysphagia and hematemesis, or vomiting of blood. Neither of these is necessarily produced by the esophagus alone, but esophageal lesions must be eliminated in the course of diagnosis. *See* ESOPHAGUS; ONCOLOGY; PATHOLOGY; *see also* CLINICAL PATHOLOGY. [E.G.ST.]

Essential oils

Odoriferous, oily products of plant origin which are distillable. The essential oils occur in the leaves, twigs, blossoms, fruit, trunk, and roots of plants. Relatively few of the great number of species of plants have been used to produce the oils of commerce.

These plants are widespread and occur even in remote geographical regions. The tropic zone produces many of them. Plants grown in areas other than those in which they were indigenous often yield oils of different and inferior composition. The function and precursors of the essential oils in plants are still not satisfactorily explained. Whether they serve as attractants, repellents, protectants, or are simply waste products is obscure.

Composition. The principal constituents of essential oils are the terpenes. Benzenoid and aliphatic compounds may also be present. Most of [illegible]

organic isothiocyanates; garlic and onion oils contain organic sulfides. Indole, skatole, and anthranilates sometimes occur in small amounts.

Of the terpenes, the monoterpenes and sesquiterpenes are the most abundant components of essential oils. In some essential oils, a terpene is the major constituent. For example, bois de rose oil contains about 85% linalool, palmarosa about 84–90% geraniol, and lemon-grass oil over 75% citral. It is frequently advantageous to isolate these major constituents from the oils for perfume and flavor applications, as well as to convert them to other important materials.

Other essential oils contain large amounts of nonterpene components. Methyl salicylate comprises about 98% of oil of sweet birch, anethole about 90% of anise oil, and eugenol as much as 95% of clove oil.

Most essential oils are exceedingly complex mixtures of terpene and nonterpene ingredients. For example, 75 identifiable compounds have been isolated from camphor oil.

Analysis. Chemical, physical, and organoleptic (odor or taste) analyses serve to determine the quality of the oils. Paper chromatography, infrared spectroscopy, and vapor fractometry have begun to play significant roles in evaluating the essential oils and perfumes. All methods possess limitations, and odor and taste must remain the final criteria, since most end uses take advantage of these qualities of the oils.

Methods of production. The most common method of separating the essential oils from plant materials is steam distillation. In order to recover the maximum amounts of oil, the plant material is either crushed or comminuted. Unique processes have been developed for almost every oil.

The flavor oils in the citrus fruits are held in tiny sacs in the rind. They are mechanically ruptured and the oils collected by decantation, centrifugation, and filtration. The oils of orange, lemon, mandarin, and bergamot are obtained commercially in this way.

Because the oils from blossoms such as jasmine, gardenia, tuberose, and violet are destroyed or lost during steam distillation, a process termed enfleurage has been used. The blossoms are placed in contact with purified fats, which become saturated with the odorous oils. Extraction with alcohol then removes the oils from the fats.

In some cases plant materials are extracted by volatile solvents such as petroleum ether and benzene. Such extractions remove not only most of the perfume materials but also waxes, coloring matter, and resinous materials. The solvents are removed, and the residue is frequently re[illegible] with alcohol to obtain oils of higher purity.

flower oils are treated in this way, as are olibanum, myrrh, opopanax, ginger, capsicum, and celery.

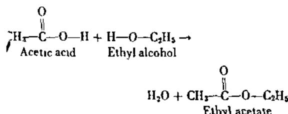
Concentration. Occasionally it is advantageous to concentrate an essential oil in order to achieve greater solubility, stability, and strength of odor. The isolation of the hydrocarbon terpenes is the chief objective, and this may be done by fractional distillation and aqueous alcoholic extraction. The final products are somewhat changed from the original oils.

Synthesis. No essential oil has been successfully reconstructed. Synthetic methyl salicylate can be differentiated by odor and taste from oil of sweet birch even though the natural oil contains 98% methyl salicylate.

Synthetic methods have made possible the commercial production of important constituents of essential oils, such as linalool, linalyl acetate, nerolidol, citral, β-ionone, geranyl acetate, nerol, and α-irone. These developments presage an i teresting future in the perfume and flavor field. See ISOPRENE; TERPENE. [E.L.S.]

Ester

The product of intermolecular dehydration involving an alcohol and an acid, according to the following equation:



Structurally, a typical ester (ethyl acetate) is thus composed of the two functional groups, keto



and *ether* (C—O—C), and the physical and chemical properties of low-molecular-weight esters reflect, in general, a blending of these two properties. With increase in molecular weight the hydrocarbon portion of the molecule overbalances the oxygen functions; thus, the higher esters are bland, water-insoluble, solid waxes. *See* WAX, ANIMAL AND VEGETABLE.

Because an extraordinarily large number of acids and alcohols are available for reaction, literally thousands of esters have been prepared and studied. These range from the simplest ($HCOOCH_3$, methyl formate) to high-molecular-weight compounds containing 30 or more carbons in both portions of the molecule.

Esterification. The process whereby esters are formed has been much studied. About 1880, N. Menschutkin found that, under controlled conditions, the rate of esterification was related to the structure of the alcohol. All the normal alcohols (primary) esterify at about the same rate, but faster than the isomeric secondary alcohols, which in turn esterify faster than the tertiary alcohols. The structure of the acid also plays a part. In an homologous series of acids, increasing complexity of structure slows down the rate of esterification.

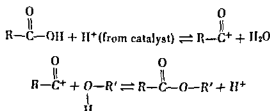
The fact that esterification involves an equilibrium was established in 1862 by M. P. E. Berthelot in his study of the ethyl alcohol–acetic acid system. If 1 mole each of acetic acid and ethyl alcohol react, it is found at equilibrium that ⅔ mole each of ethyl acetate and water are present at room temperature, along with ⅓ mole each of alcohol and acid. This can be applied to the equilibrium equation:

$$K_E = \frac{[CH_3COOC_2H_5] \times [H_2O]}{[CH_3COOH] \times [C_2H_5OH]}$$

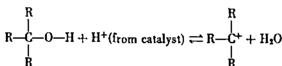
where K_E is the equilibrium constant and the square brackets signify concentrations in moles per liter of the enclosed reagent. This gives for K_E the value 4. Unless the temperature is deliberately changed, this value is fixed for ethyl alcohol and acetic acid (different alcohol-acid systems have different though characteristic equilibrium constants); indeed, regardless of the starting concentrations of acid and alcohol, the value 4 is maintained.

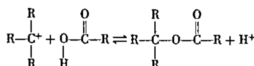
The rate of esterification is subject to catalysis; the usual method of speeding the attainment of equilibrium consists of the addition of about 3% (by weight) of dry hydrogen chloride, concentrated sulfuric acid, or boron trifluoride. If the reactants are heated in presence of a catalyst, equilibrium can usually be attained in 2–3 hours except for those cases involving an aromatic acid carrying substituents in both ortho positions with respect to the carboxyl. V. Meyer (1894) showed that such acids were so hindered in their reaction with alcohols that direct esterification was impracticable. A generally successful method for esterifying such hindered acids is now available (M. S. Newman, 1941), and consists of first dissolving the substituted aromatic acid in 100% sulfuric acid, and then pouring the solution into the required alcohol. There is a good yield of ester in only a few minutes.

The mechanism of direct esterification has been much studied. The use of isotopic oxygen (O^{18}) shows that in the reaction of an acid with an alcohol of primary or secondary type, the ester oxygen comes from the alcohol and the acid oxygen goes to form water. Moreover, in an ordinary acid-catalyzed esterification, the rate of reaction is dependent upon the concentrations of both the carboxylic acid and the alcohol. These observations are accommodated by the following mechanistic picture:



In the case of tertiary alcohols, isotopic oxygen studies show that the ester oxygen comes from the carboxylic acid, and the hydroxyl from the alcohol goes to form water, implying the following modified picture:



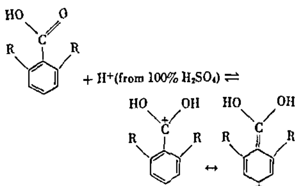


The Newman method for hindered aromatic acids is not effective for unhindered acids. It involves the following mechanism in which 100% sulfuric acid is used:

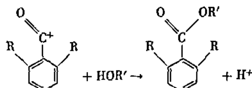
Some methyl and ethyl esters*

Name	Formula	Melting point, °C	Boiling point, °C
Methyl esters:			
Abietate	$C_{19}H_{29}COOCH_3$	liq.	360–365 (decomposes)
Acetate	CH_3COOCH_3	−98 7	57.3
Acetoacetate	$CH_3COCH_2COOCH_3$	−80	169–170
Acrylate	$CH_2{=}CHCOOCH_3$		80.3 (polymerizes)
Adipate	$[(CH_2)_2COOCH_3]_2$	10–11	$115^{11\ mm}$
Benzoate	$C_6H_5COOCH_3$	−12.4	199.5
Butyrate	$CH_3(CH_2)_2COOCH_3$	−84.8	102.8
Caproate	$CH_3(CH_2)_4COOCH_3$	−71.0	151.3
Citrate	$C_3H_5O(COOCH_3)_3$	78–79	283 (slight decomposition)
Formate	$HCOOCH_3$	−99	31 8
Malonate	$CH_2(COOCH_3)_2$	−62	180–181
Myristate	$CH_3(CH_2)_{12}COOCH_3$	18–19	$295^{751\ mm}$
Oxalate	$(COOCH_3)_2$	54	163.3
Pelargonate	$CH_3(CH_2)_7COOCH_3$		$213–214^{757\ mm}$
o-Phthalate	$o\text{-}C_6H_4(COOCH_3)_2$		$280^{724\ mm}$
m-Phthalate (isophthalate)	$m\text{-}C_6H_4(COOCH_3)_2$	67–68	
p-Phthalate (terephthalate)	$p\text{-}C_6H_4(COOCH_3)_2$	140	>300 (sublimes)
Propionate	$CH_3CH_2COOCH_3$	−87 5	79.7
Salicylate	$o\text{-}HO{-}C_6H_4COOCH_3$	−8 6	222.9
Succinate	$[CH_2COOCH_3]_2$	18 2	195 9
Tartrate	$(CH_3)_2C_4H_4O_6$	61.5	280
Ethyl esters:			
Abietate	$C_{19}H_{29}COOC_2H_5$	Dec. >280	$200^{4\ mm}$
Acetate	$CH_3COOC_2H_5$	−83.6	77.2
Acetoacetate	$CH_3COCH_2COOC_2H_5$		198
Acrylate	$CH_2{=}CHCOOC_2H_5$		100.1 (polymerizes)
Adipate	$[(CH_2)_2COOC_2H_5]_2$	−19.8	$138\ 8^{15\ mm}$
Benzoate	$C_6H_5COOC_2H_5$	−34.7	212.4
Butyrate	$CH_3(CH_2)_2COOC_2H_5$	−100.8	121 6
Caproate	$CH_3(CH_2)_4COOC_2H_5$	−67.5	167 9
Citrate	$C_3H_5O(COOC_2H_5)_3$		294
Formate	$HCOOC_2H_5$	−79.4	54.2
Glutarate	$(CH_2)_3(COOC_2H_5)_2$	−23.8	233.7
Malonate	$CH_2(COOC_2H_5)_2$	−51.5	199.3
Myristate	$CH_3(CH_2)_{12}COOC_2H_5$	10.5–11.5	295
Oxalate	$(COOC_2H_5)_2$	−40.6	185.4
Pelargonate	$CH_3(CH_2)_7COOC_2H_5$	−36 8	227
o-Phthalate	$o\text{-}C_6H_4(COOC_2H_5)_2$		289.5
m-Phthalate (isophthalate)	$m\text{-}C_6H_4(COOC_2H_5)_2$	11.5	302
p-Phthalate (terephthalate)	$p\text{-}C_6H_4(COOC_2H_5)_2$	43–44	302
Propionate	$CH_3CH_2COOC_2H_5$	−73 9	99.1
Salicylate	$o\text{-}HO{-}C_6H_4{-}COOC_2H_5$	1.3	233–234
Succinate	$[CH_2COOC_2H_5]_2$	−21.3	217
Tartrate (+)	$(C_2H_5)_2C_4H_4O_6$	17	280
o-Toluate	$o\text{-}CH_3C_6H_4COOC_2H_5$	<−10	227
m-Toluate	$m\text{-}CH_3C_6H_4COOC_2H_5$		$231^{750\ mm}$
p-Toluate	$p\text{-}CH_3C_6H_4COOC_2H_5$		235.5

* Based on N. A. Lange (ed.), *Handbook of Chemistry*, 9th ed., McGraw-Hill, 1956



For stabilizing resonance to occur in the latter ion, coplanarity of ring and hydroxyls is necessary. This is possible with unsubstituted benzoic acids, *but is not possible with di-ortho-substituted acids;* water is therefore eliminated, giving the acyl ion which can react readily with an alcohol:

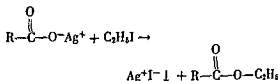


Hydrolysis. The splitting of esters in such a way as to regenerate the parent acid and alcohol is an example of hydrolysis. It is frequently important, especially in dealing with naturally occurring es- s such as those found in animal and vegetable *fats, oils, and waxes. In the presence of dilute mineral* acid, hydrolysis of an ester is the reverse of acid-catalyzed esterification; an excess of water is used to ensure complete splitting, and the reaction is carried out at elevated temperatures to speed up the process. Often alcohol is added to solubilize the reactants. Esters formed from glycerol and long-chain carboxylic acids (fats and oils), from long-chain acids and long-chain alcohols (waxes), and simple esters of mono-, di-, or polycarboxylic acids with primary, secondary, or tertiary alcohols, are hydrolyzeable under acid conditions, using dilute hydrochloric or sulfuric acids, or Twitchell's reagent (prepared from benzene or naphthalene, oleic acid, and concentrated sulfuric acid). However, esters of di-ortho-substituted aromatic carboxylic acids (for example, 2,6-dimethylbenzoic acid) are hindered with respect to hydrolysis, and must be treated according to the Newman technique, which involves first solution in 100% sulfuric acid, and then addition to excess cold water. This technique, which is ineffective for the hydrolysis of nonhindered esters, is presumed to involve the *intermediacy of the resonating acyl ion:*

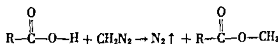
Hydrolysis under basic conditions, as in the manufacture of soaps, is effective for esters of primary and secondary alcohols, but not tertiary, which are split only by acid hydrolysis.

Esters may be made by several methods not involving direct esterification. Chief among these are the following:

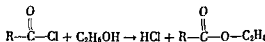
1. Interaction of a metallic salt (usually silver) of an acid, with an alkyl halide,



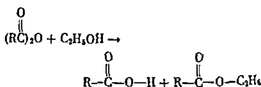
2. Reaction of diazomethane with an acid, for the specific formation of a methyl ester,



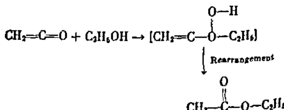
3. Elimination of hydrogen halide from an acyl halide and a primary or secondary alcohol (*see* Acid halide),

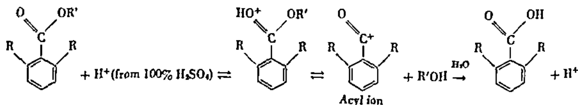


4. The reaction between an acid anhydride and an alcohol (*see* Acid anhydride),

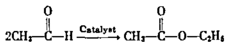


5. With ketene, esters of acetic acid are produced, through the following mechanism,





6. By the Tischenko reaction, which involves a simultaneous oxidation and reduction of two aldehyde molecules, catalyzed by aluminum ethoxide,



$$2CH_3{-}\overset{\overset{\displaystyle O}{\|}}{C}{-}H \xrightarrow{\text{Catalyst}} CH_3{-}\overset{\overset{\displaystyle O}{\|}}{C}{-}O{-}C_2H_5$$

The table lists some representative methyl and ethyl esters.

Reactions and uses. Esters undergo a variety of reactions, many of which are industrially important. With ammonia, the ester linkage is split, yielding the alcohol and the amide (*see* AMIDE, ACID). Acid or basic hydrolysis yields the alcohol and the free carboxylic acid or its salt, respectively; basic hydrolysis applied to animal or vegetable fats and oils produces metallic salts of long-chain organic acids along with the alcohol, glycerol. *See* SOAP AND DETERGENT.

Catalytic reduction of esters can be effected at elevated temperature (250°C) and pressure (15–20 atm), using molecular hydrogen and a copper chromite catalyst; this furnishes a convenient means for the preparation of long-chain mono- or dihydroxy alcohols from esters of the corresponding mono- or dicarboxylic acids. Thus, diethyl succinate ($C_2H_5OOC{-}CH_2CH_2{-}COOC_2H_5$) is reduced to form ethyl alcohol (C_2H_5OH) and butylene glycol ($HO{-}CH_2CH_2CH_2CH_2{-}OH$). For laboratory reductions, either sodium dissolving in alcohol (Bouveault-Blanc method), or lithium aluminum hydride is preferred.

Esters usually react well with Grignard reagents to yield tertiary alcohols. *See* GRIGNARD REACTION.

Low-molecular-weight esters are liquids, and are much used as solvents, particularly the acetates of methyl, ethyl, and butyl alcohols. Esters of cellulose (cellulose triacetate) are used in photographic film, and as a textile fiber (acetate rayon). Cellulose acetatepropionate and cellulose acetatebutyrate have become important as thermoplastic materials. Cellulose nitrate, containing 10.5–11% nitrogen, is called celluloid pyroxylin; with alcohol and camphor (a plasticizer), it forms celluloid. Dynamite cotton is cellulose nitrate of 11.5–12.3% nitrogen content, and gun cotton is cellulose nitrate of 12.5–13.5% nitrogen. Cordite and ballistite are made from the latter, which is plasticized with glyceryl trinitrate (nitroglycerine). Dimethyl and diethyl sulfates (esters of sulfuric acid) are excellent agents for alkylating organic molecules that contain labile hydrogen atoms, for example, starch and cellulose. *See* EXPLOSION AND EXPLOSIVE.

Esters of unsaturated acids, for example, acrylic or methacrylic acid, are unstable and polymerize rapidly, yielding resins; thus, methyl methacrylate yields a polymethyl methacrylate resin (lucite). Analogously, esters of unsaturated alcohols are unstable and readily react with themselves; thus, vinyl acetate polymerizes to polyvinyl acetate. The polyester resins known as glyptals result from the polyesterification of glycerol with phthalic anhydride; the process can be controlled to yield either a fusible or an infusible resin. When the polyesterification is carried out in the presence of a long-chain, unsaturated acid of the drying oil type, the oxidative polymerization of the latter is superimposed upon the polyesterification, resulting in hard, synthetic, weather-resistant enamels, suitable for automobile finishes. Polyesterification of ethylene glycol with terephthalic acid results in a polyester fiber. If the material is formed in sheets, it is a useful photographic film.

Many low-molecular-weight esters have characteristic, fruitlike odors: banana (isoamyl acetate), rum (isobutyl propionate), and pineapple (butyl butyrate). These esters are used to some extent in compounding synthetic flavors and perfumes. *See* ALCOHOL; CARBOXYLIC ACID; DRYING OIL; FAT AND OIL, NONEDIBLE; POLYESTER RESINS; SOLVENT.

[E.B.R.]

Esterification

Any reaction in which at least one of the products is an ester. An ester is a compound (RCOOR′) in which an acyl group (RCO—) is attached to an alkoxy group (R′O—). Here, R and R′ usually represent hydrocarbon radicals. If an inorganic acid, such as nitric, sulfuric, or phosphoric acid is used, the acyl group is replaced by the nitrate ($—NO_2$), acid sulfate ($—SO_3H$), or the diacid phosphate ($—PO_3H_2$) group. *See* ESTER.

There are many routes to the formation of esters. Some of the more important reactions for preparing esters take place between the following pairs of compounds: (1) an acid and an alcohol, (2) an acid anhydride and an alcohol, (3) an acid chloride and an alcohol, (4) an acid and an unsaturated hydrocarbon such as an olefin or an acetylene, (5) an ester and an alcohol, (6) an ester and an acid, and (7) two different esters. This article treats esterification in only a limited sense —the reaction between a carboxylic acid (RCOOH) and an alcohol (R′OH) to give the ester and water. For discussions of reactions of an ester with an alcohol, an acid, or another ester, *see* TRANSESTERIFICATION.

Esterification reactions are generally reversible and accompanied by relatively small heat effects of the order of a few kilocalories per mole of ester. Although the reactions generally take place in a single liquid phase in the presence of a catalyst, a limited number of esters have been prepared by passing the reactant vapors over a solid catalyst. In the presence of a catalyst, the reaction is commonly conducted at a temperature of about 100°C; in the absence of a catalyst, a temperature of about 250°C is used to give a reasonable reaction rate. The pressure at which the reaction is conducted is determined only by the volatility of the components of the system. It is usually atmospheric pressure. In order to produce most esters economically, some means must be provided for completing the reaction by removing one or more of the products.

A few of the important commercial [illegible] of esterification follow. Dibutyl phthalate is

pared from phthalic anhydride and butanol in a stepwise reaction to form first the monoester and then the diester; cellulose acetate from purified α-cellulose and a mixture of acetic anhydride and acetic acid; alkyd resins from phthalic anhydride, unsaturated fatty acids, and glycerol; nitroglycerine (glycerol trinitrate) from glycerol and the proper mixture of nitric acid and sulfuric acid; ethyl acetate from ethyl alcohol and acetic acid.

Reaction equilibria. The fraction of the limiting reactant which is converted at equilibrium is significantly less than unity for reactions between the lower-molecular-weight acids and alcohols. In the case of the higher-molecular-weight fatty acids, such as stearic, an equilibrium conversion of essentially 1 is attained. Defining the concentration equilibrium constant as

$$K_c = \frac{[\text{ester}]\,[\text{water}]}{[\text{acid}]\,[\text{alcohol}]}$$

where the brackets designate the concentration at equilibrium; values of K_c for the low-molecular-weight esters range between about 1 and 10 at 155°C. Because of activity effects, these equilibrium constants tend to increase with an increase in the initial acid to alcohol ratio, and to decrease slightly if water is present initially.

Reaction kinetics. In the absence of an added catalyst, rates of esterification commonly have been found to be proportional to the first power of the alcohol concentration and to the second power of the acid concentration. In the presence of a strong acid catalyst, the rate commonly is proportional to the first power of the catalyst concentration, the first power of the esterifying acid concentration, and the first power of the alcohol concentration. With a large excess of alcohol, the rates of a large number of esterification reactions have been found to obey the Godschmidt equation

$$\text{Rate} = \frac{k[\text{acid}]\,[\text{catalyst}]\,r}{r + [\text{water}]}$$

where r is a constant characteristic of the relative affinity of the alcohol and water for hydrogen ions, and k is the specific rate constant.

Without any added catalyst, the reaction rate approximately doubles with each 10°C rise in temperature. The reaction rate may be increased by a large factor in the presence of a catalyst, accompanied by a decrease in the temperature coefficient to a value less than half that for the uncatalyzed reaction.

Catalytic agents. The effect of adding a catalyst is to increase the reaction rate with little effect on the equilibrium constant. All good esterification catalysts have been compounds which are strongly acid in nature, either in the Bronsted or the Lewis sense. The most important commercial catalysts are hydrochloric and sulfuric acid, with sulfuric acid being preferred because of its lower cost and lower corrosive effect on metals. Other strong acids which have found wide use in the laboratory are perchloric acid, and aromatic phosphonic and sulfonic acids, such as *p*-toluenesulfonic acid. Strongly acidic ion-exchange resins are finding favor because of the ease of separating the catalyst from the reaction products and the decreased corrosion problem. A number of salts have shown strong catalytic activity. Foremost among these are the fluorides of boron and silicon, and the chlorides of zinc, aluminum, and tin. Also, various soaps and sulfonic acid salts of calcium, magnesium, manganese, tin, and cobalt have been utilized to some degree. The oxides of silicon, titanium, and thorium are particularly effective catalysts for the reaction in the vapor phase.

Effects of molecular structure. Classical studies by N. Menschutkin with acetic acid and a variety of alcohols at 155°C show a slight increase in the specific rate constant and a slight decrease in the equilibrium constant with an increase in molecular weight for the low-molecular-weight aliphatic alcohols. When the number of carbon atoms remains fixed, the specific rate constant tends to decrease with chain branching, proximity of the branching to the —OH group, location of the —OH group on a secondary or tertiary carbon atom, presence of unsaturation in the chain, and introduction of an aromatic nucleus. Similar studies at 155°C with isobutyl alcohol and a variety of acids give corresponding results as to the effect of molecular weight, branching, and unsaturation on the specific rate constant. The effects on the equilibrium constant are generally reversed from those obtained by changing the alcohol. Larger equilibrium constants are obtained with greater chain length, increased substitution, and for the aromatic acids. The purely aromatic acids esterify very slowly

Reaction mechanism. The bonds broken in esterification are the carbonyl-oxygen of the acid and the hydrogen-oxygen of the alcohol. A bond is formed between the carbonyl carbon atom of the acid and the oxygen of the alcohol. The basis for this is the stronger electronegativity of oxygen relative to carbon; a positive charge is believed to be induced on the carbonyl carbon atom by the carbonyl oxygen atom. This effect is enhanced by the presence of a hydrogen ion in the vicinity of the carbonyl oxygen. Electron-attracting substituent groups in the hydrocarbon radical of the acid increase the specific rate constant.

Industrial practice. Industrial esterifications are usually accompanied by the removal of at least one of the reaction products in order to carry the reaction to completion. Reduced pressure, sparging with an inert gas such as carbon dioxide, and azeotropic distillation all are used. Frequently, the reactor and distillation column are combined. The operation may be batch, semibatch, or continuous. The specific conditions of feed rates, catalyst concentration, mole ratio of the alcohol to acid,

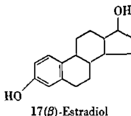
entrainer (if used), and the number of plates in the distillation column will depend upon the ester being prepared. *See* SUBSTITUTION REACTION; UNIT PROCESSES. [J.M.WO.]

Bibliography: H. A. Goldsmith, Polyhydric alcohol esters of fatty acids, ***Chem. Revs.***, 33:257–349, 1943; P. H. Groggins, ***Unit Processes in Organic Synthesis***, 5th ed., 1958; C. K. Ingold, ***Structure and Mechanism in Organic Chemistry***, 1953.

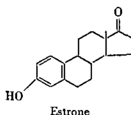
Estrogen

A substance that maintains the secondary sex characters and organs, such as mammary glands, uterus, vagina, and fallopian tubes, of mammalian females. Naturally occurring substances with this activity are steroid hormones. The principal estrogenic hormone substances are 17(β)-estradiol, estrone, and estriol. These compounds have a ring A phenolic group. They are produced and secreted directly into the blood stream by the ovary, testis, adrenal, and placenta of pregnancy. Two other naturally occurring estrogenic hormones, equilin and equilenin, have been obtained only from the urine of pregnant mares and are apparently peculiar to that species. Stilbestrol, a synthetic compound with considerable estrogenic activity, has been used extensively in medical practice.

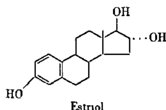
Ring A phenolic group estrogens. 17(β)-Estradiol is the most active of the estrogens known to be produced by mammalian tissues and is interconvertible with estrone. An enzyme present in liver, placenta, and other tissues catalyzes this oxidation-



17(β)-Estradiol



Estrone

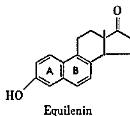


Estriol

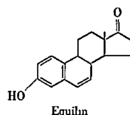
reduction. Estriol is considered to be derived from estrone.

Ring B unsaturated estrogens. Equilenin and equilin are naturally occurring representatives of ring B unsaturated estrogens. Equilenin also has been detected in human adrenal cancer tissue from

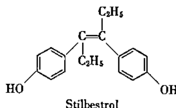
tion, and a higher activity by oral administration.



Equilenin



Equilin



Stilbestrol

Biosynthesis of estrogens. This involves a series of reactions from androgens such as testosterone and Δ^4-androstene-3,17-dione.

This mechanism probably accounts for most of 17(β)-estradiol and estrone, the two most active, naturally occurring estrogens. The transformation of androgens to estrogens involves the removal of carbon 19 and two hydrogen atoms at carbons 1 and 2. The loss of the carbon atom most likely occurs by way of hydroxylation at carbon 19, and then successive oxidative reactions to carbonyl, carboxylic acid functional groups, and finally evolution of CO_2. The mechanism involving the loss of

of the elements of water. Biosynthesis of estrogens from androgens is not limited to gonads, adrenals, and placenta, since the transformation is demonstrable in women devoid of these glands. The liver is probably the most important additional biosynthetic site.

Testosterone

17(β)-Estradiol

Estrone

Biosynthesis of estrogens

6(β)-Hydroxy-17(β)-estradiol

18-Hydroxyestrone

2-Methoxyestrone

17(β)-Estradiol

Estrone

17(α)-Estradiol

Estriol

Epiestriol

Metabolic breakdown of estrogens

Equilenin and equilin appear to be biosynthesized by a pathway different from that for 17(β)-estradiol and estrone. Whereas the bulk of 17(β)-estradiol and estrone is derived from cholesterol by way of progesterone and androgens, the ring B unsaturated estrogens, such as equilenin and equilin, appear to be derived from a noncholesterol source. This is probably a dietary C_9 compound and a C_{10} which could arise from acetate. The possibility exists that ring B unsaturated estrogens may be converted to 17(β)-estradiol and estrone.

Catabolism of estrogens. The metabolic breakdown of estrogens involves various modifications at carbon atoms 2, 6, 16, and 18. These over-all reactions are illustrated in the series of equations. [R.I.D.]

Estrus

The period in mammals during which the female ovulates and is receptive to mating. It is commonly referred to as rut or heat. From one estrus period to the next there occurs a series of changes, particularly in the ovary, uterus, and vagina, termed the estrous cycle. With reference to the ovary, the cycle can be divided into a follicular phase, during which the Graafian follicles are ripening, and a luteal phase, during which the corpora lutea develop in the ovulated follicles. During these two phases mainly estrogen and progesterone, respectively, are secreted, and these hormones control the uterine and vaginal changes. The beginning of the follicular phase is termed proestrus and the luteal phase metestrus. Following the latter, there is a period of relatively little change, termed diestrus. In species in which the latter is prolonged, it is termed anestrus.

During proestrus, the endometrial lining of both the uterus and vagina thickens. The vagina regresses in metestrus, and vaginal washings exhibit mainly cornified epithelial cells. At the same time uterine glands continue to enlarge and become highly secretory. This is the stage during which implantation of the embryo would occur if the previously ovulated eggs had been fertilized. Toward the end of this period, in the absence of fertilization, the uterine lining regresses, gradually in most mammals, but precipitously in primates. In the latter the regression involves considerable tissue loss and bleeding and is termed menstruation. The menstrual cycle of primates, typically about 28 days, is essentially an estrous cycle with ovulation, or estrus, usually occurring during the middle week but with mating not limited to this time.

Some mammals such as the armadillo, bat, deer, wolf, shrew, and ferret are monestrous, having a single cycle per year. Most are polyestrous, but considerable variations occur in different species. In dogs, there are 2–3 cycles per year Cats cycle about every two weeks during breeding seasons. Cows, horses. and sheep illustrate seasonal polyestrus with 2½- to 3-week cycles, while laboratory rats and mice maintain a 5- to 6-day schedule throughout the year. In the rabbit, there is persistent estrus, and ovulation occurs in response to the stimulus of coitus. This mating-induced ovulation is known to occur also in the cat and ferret during their more restricted period of estrus, and probably occurs in a number of other mammals.

The regulation of the estrous cycle, and of breeding seasons, is brought about by feedback-type hormonal interactions between the ovary and pituitary gland, influenced in many cases by environmental factors, such as length of day. Thus, for example, if additional illumination is provided in the fall, ferrets can start breeding during the winter, instead of the spring; and if the sequence of night and day is changed, mice, which ordinarily mate and ovulate during the night, will adapt to the new conditions. *See* OVUM; PITUITARY GLAND; REPRODUCTION, ANIMAL. [A.TY.]

Estuarine oceanography

The study of the chemical, physical, biological, and geological properties of estuaries. An estuary is a semienclosed coastal body of water having a free connection with the open sea and within which sea water is measurably diluted with fresh water. Because of similarities in the controlling physical and chemical processes and in the techniques employed in their study, the field of estuarine oceanography is often considered to include investigations of certain coastal water bodies which are not, by strict definition, estuaries. Bays and lagoons in which evaporation is equal to or exceeds fresh-water inflow, so that the salt content equals or exceeds that of the adjacent open sea, are generally included in the field of study.

Classification. Estuaries may be classified according to several different criteria; no one system is universally used. A broad classification separates true estuaries, called positive estuaries, in which fresh-water inflow exceeds evaporation, from those embayments, called inverse estuaries, in which evaporation exceeds fresh-water inflow. In this broad classification neutral estuaries are those embayments in which neither evaporation nor fresh-water inflow dominates.

Estuaries are also classified on the basis of geomorphological structure. Coastal plain estuaries are formed by the drowning of river mouths and the lower reaches of river valleys, either from subsidence or from a rise in sea level. This type of estuary is usually an elongate indenture of the coastline with a river flowing into the upper end. The resulting body of water is typically relatively shallow, sometimes with an irregular (dendritic) shoreline. The eastern coast of North America is characterized by estuaries of this class, such as Delaware Bay, the lower Hudson River, the lower Savannah River, and Chesapeake Bay with its tributary estuaries. These systems are true estuaries; however, in some cases, either nature or man has diverted the inflowing river to a new outlet to the sea, leaving an embayment of the coastal-plain type in which evaporation exceeds fresh-water inflow. *See* COASTAL LANDFORMS.

Glacially cut fiords of the Norwegian and Canadian Pacific coasts are representatives of a second geomorphological class of estuaries, called deep-basin estuaries. These elongated indentures of the coastline contain a relatively deep basin with a shallow sill at the mouth. *See* FIORD

A third group of estuaries results from the development of an offshore bar on a shoreline of low relief. These bar-built estuaries have a rather narrow connection with the open sea. They are elongated parallel to the coast, rather than being indentures into the coastline. Many bodies of water formed by offshore bars are not true estuaries, as they frequently exhibit an excess of evaporation over fresh-water inflow.

An essential feature of an estuary is the intermixing of sea water with fresh water derived from land drainage This intermixing produces a variation, both vertically and horizontally, in the salt content of estuarine waters. Oceanographers use the term salinity to designate the concentration of dissolved solids in oceanic and estuarine waters. Salinity is generally expressed in terms of grams of dissolved material per thousand grams of sea water, that is, in parts per thousand (‰). Salinities in estuaries range from near zero at the head to approximately 30‰ at the mouth. There is also generally an increase in salinity with depth. *See* SEA WATER.

Physical structure and circulation. Estuarine circulation refers to the gross features of the pattern of estuarine water movements. Water movements in estuaries result primarily from the interaction of three processes: tides, river flow, and wind. The first two of these are usually dominant, and control the distribution of salinities, though in bar-built estuaries wind-induced motion is often the most important. The physical structure of estuarine waters refers to the vertical and horizontal distribution of density. Since variations in density are related primarily to variations in salt content, any treatment of estuarine circulation logically includes consideration of the physical structure.

Coastal plain estuaries. Estuaries of this type have a physical structure and circulation pattern controlled primarily by the relative magnitude of oscillatory tidal movements and motion induced by river inflow The oscillatory tidal velocities moving over the shallow bottom of the estuary produce turbulent mixing of the river inflow and the salt waters from the sea. The lower density of fresh water as compared to that of sea water produces a stable vertical stratification which resists vertical mixing. The relative magnitudes of the tidal flow and the river inflow are thus an important control on the physical structure of the estuary.

Salt-wedge estuary. In coastal-plain estuaries in which the river flow is large compared to the tidal flow, the sea water enters as a salt-water wedge along the bottom. If no frictional drag existed between the denser salt-water wedge and the lighter river water, the wedge would extend upstream to the point where the river bottom intersected mean sea level. The river inflow would move seaward over the stationary wedge as a thin fresh-water layer. In the real fluid, however, some frictional drag always exists. The extent of intrusion of the wedge up the estuary depends upon the magnitude of this frictional drag, which in turn depends upon the relative velocities in the upper, seaward-flowing layer and in the wedge. Thus the volume rate of flow of the river water which moves seaward on top of the salt-water wedge controls the position of the wedge. Under conditions of high river flow, the wedge extends only a short distance into the estuary, while for low river flows the wedge extends many miles upstream. The Mississippi River is an example of this salt-wedge estuary (Fig. 1). When the flow in the Mississippi is low, the undiluted salt wedge extends upstream for over 100 miles. Under conditions of high river flow the salt wedge extends only a mile or so above the mouth of the river.

Comparatively little mixing occurs at the interface between the seaward-flowing upper layer and the salt wedge; hence the salinity throughout the wedge is nearly that of full sea water. At the upper boundary of the wedge, unstable interfacial waves form and break into the upper layers, producing a slowly increasing salt content in these layers as they move seaward. Even so, a very sharp vertical salinity gradient exists between the upper layer of relatively low salinity and the salt wedge. For a given river flow the horizontal position of the wedge and its vertical dimension remain stationary. The loss of salt water from the wedge to the upper layer must be compensated for by flow of water into the wedge from the sea. Since this loss of salt water to the upper layer takes place all along the upper boundary of the wedge, there must be, in order to maintain the position and shape of the



Fig 1. Freighter stirring up clear Gulf water in Southwest Pass of Mississippi River. Muddy river water flows seaward over the denser, upstream flow of Gulf water. (U.S. Corps of Engineers)

wedge, a flow directed toward the upstream tip of the wedge at all positions within the wedge. Thus the circulation pattern in this salt-wedge estuary involves a seaward-flowing upper layer riding over the landward-directed flow in the salt wedge. The compensating flow in the wedge is small compared to the seaward flow of the low-salinity surface layers.

Partially mixed estuaries. In coastal-plain estuaries in which tidal movements are large compared to the river inflow, vertical mixing is sufficiently strong to destroy the sharp boundaries separating the salt wedge from the upper layer, and the wedge ceases to exist as a readily identifiable feature. There still exists a transition layer of relatively rapid increase in salinity with depth, called the halocline, which separates the lower-salinity surface layers from the higher-salinity deeper water The salinity in both the surface layers and the bottom layers decreases steadily from the mouth to the head of the estuary. Most of the estuaries along the eastern coast of the United States fall into this class of partially mixed estuaries. *See* HALOCLINE.

The oscillatory tidal currents produce the most obvious motion in these partially mixed estuaries. Superimposed on the tidal currents there is a net circulation pattern, with a net seaward flow in the surface layers, and a net flow directed from the mouth toward the head of the estuary in the deeper layers. There is also a small net vertical motion directed from the deeper layers to the surface layers. The volume of water flowing toward the head of the estuary per unit time decreases as one proceeds from the mouth to the head of the estuary, since water is being transferred through vertical motion from these deeper layers to the surface layers. Hence, the volume rate of seaward flow in the surface layers increases as one proceeds from the head toward the mouth of the estuary.

The flows involved in this net circulation pattern in partially mixed estuaries are many times larger than the fresh-water inflow. This results from the potential energy gained through the tidal mixing. The center of gravity of a column of water, the upper half of which is fresh and the lower half salt water, will be raised when the two halves are mixed together. Thus kinetic energy of the oscillatory tidal motion is transferred through the disorganized motion of turbulence to increased potential energy of the water column, which in turn provides for increased kinetic energy of the net circulation pattern.

Vertically homogeneous estuaries. In this third group of coastal-plain estuaries the tidal movements are very large compared to the motion induced by the inflowing fresh water. The mixing induced by the tidal motion completely overcomes the stability resulting from the fresh-water inflow, producing uniform salinity from surface to bottom The salinity decreases from the mouth to the head of the estuary. Also, in relatively wide estuaries the salinity on the right side (looking toward the mouth) will be lower than the salinity on the left side, as a result of the effects of the earth's rotation (Fig. 2). *See* CORIOLIS ACCELERATION AND FORCE.

The circulation pattern in a vertically homogeneous estuary shows no variation in water movement with depth. In relatively wide estuaries, net seaward flow occurs along the right side (looking toward the mouth of the estuary), and a net motion directed toward the head of the estuary occurs on the left side. A laterally directed flow carries water from the left side of the estuary to the right, and large-scale horizontal mixing occurs between the counterflows on the two sides of the estuary.

In narrow, vertically homogeneous estuaries tidal mixing may be sufficient also to destroy the lateral salinity gradient. Such an estuary is said to be sectionally homogeneous. The only variation in salinity is the decrease from the mouth to the head of the estuary. The net circulation pattern is quite simple, being a slow seaward movement at all depths.

Deep-basin estuary. The deep-basin, or fiord, type of estuary has a physical structure and circulation pattern similar to those which would be expected if the solid bottom in a partially mixed coastal-plain estuary were replaced by a deep basin filled with sea water. The significant circulation pattern and variations in salinity are restricted to the upper 20 meters or so. Below this depth the salinities are nearly constant and approximately the same as the salinity of the corresponding depths, above the depth of the sill, in the adjacent open sea. Many fiords have rivers flowing into the upper ends. In such estuaries the salinity of the surface layers is low at the head of the fiord and increases steadily toward the mouth. This surface layer is several meters deep as a result of tidal and

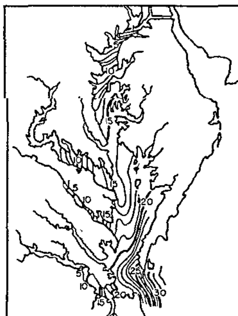


Fig 2. Typical surface salinity distribution in Chesapeake Bay. (*From D. W. Pritchard, Estuarine Hydrography, in H. E. Landsberg, ed, Advances in Geophysics, vol. 1, Academic Press, 1952*)

wind mixing. Below the surface layer a halocline marks the transition layer, in which the salinity increases with depth from the low value found in the surface layer to the high salinity of the deep basin.

Superimposed on the oscillatory tidal motion there exists a net circulation pattern in which the water in the transition zone flows toward the head of the fiord. There is also a slow net vertical movement from the transition zone to the surface layers, such that the volume rate of flow seaward in the surface layers increases in the seaward direction, while the volume rate of flow toward the head of the fiord in the transition layer decreases along the direction of flow. Partial mixing takes place between the surface layer and the transition zone. There is very little motion in the deep waters of the basin. For a discussion of the deep circulation in fiords (marginal seas) *see* OCEAN CURRENTS.

Bar-built estuary. This class of estuary exhibits no well-established pattern of salinity distribution and circulation. Except at the narrow inlets connecting the sound (as this type of estuary is frequently called) with the open sea, these estuaries are usually so shallow that wind mixing produces vertical homogeneity. Tidal current velocities through the inlets are relatively large. Because of the narrowness of the inlet, however, the total volume of water which flows in and out with the tide is relatively small, and the tidal rise and fall, as well as the tidal currents, are greatly reduced within the estuary. In the inlets proper there frequently occur a salinity structure and circulation pattern characteristic of a partially mixed estuary. Within the sound the wind produces the most significant water movements.

There must occur a net movement through and out of the estuary just sufficient to remove the fresh water added to the estuary by river inflow, runoff and direct precipitation. Because of the relatively large cross-sectional area, this net flow is usually not directly measurable within the bar-built estuary [illegible] flushing de-notes the composite processes whereby the water *within* the estuary is renewed, both through fresh-water inflow and through exchange with the waters of the adjacent coastal area.

Renewal and flushing time. The renewal time is the time required to replace a stated percentage of the water within the estuary with new water. Thus the 99% renewal time is the time required to replace 99% of the water in the estuary with new water, both through fresh-water inflow and through exchange with sea water.

Theoretically, an infinite time is required to renew all the water in the estuary, the mathematical expression for the process having an exponential form. Thus, the average retention time is not the time required to replace 50% of the water in the estuary, but rather the time required to replace all but $1/e$ of the total volume, or approximately 63% of the water in the estuary. This average retention time has also been called the flushing time of an estuary.

The 50% renewal time, or the time interval required to replace 50% of the water present in the estuary at any time with new water, is called the half-life of the estuary, by analogy with the processes of radioactive decay.

The concepts of renewal time and flushing time can be applied to segments of the estuary as well as to its total volume. When used in this sense, the source of new water for the exchange with any given segment is the next most seaward segment of the estuary, rather than the sea water at the mouth of the estuary. Theoretical treatment of estuarine flushing has been successful only for the relatively simple, sectionally homogeneous estuary.

Interest in estuarine flushing has been stimulated by the problems associated with the introduction of pollutants into the estuarine environment, both as a result of man's activity and through natural processes. The over-all problem involves the dilution and dispersal of the contaminant within the estuary, as well as its ultimate discharge from the estuary through processes of exchange. Thus estuarine flushing as discussed above includes only part of the problem. There has been a recent tendency to include within the area of study of estuarine flushing the processes of dilution and dispersion of a contaminant within the estuary.

Distribution of pollutants. When a contaminant, such as an industrial or municipal waste, is introduced into an estuary, an initial dilution occurs which is related to the manner of introduction and to the physical properties of the contaminant. The resulting contaminated volume then participates in the oscillatory tidal motion and in the net circulation pattern of the estuary, and is dispersed through the process of turbulent diffusion. Finally, the exchange of waters within the estuary with the open sea leads to the transport of the contaminant out of the estuary.

If the contaminant is in the form of a suspension of particulate material, then the added processes of settling and resuspension must be included. Waste material initially in solution may be adsorbed on the natural suspended silt, or may be taken up by the flora and fauna of the estuary, introducing added processes which must be considered.

The initial density of the introduced contaminant, relative to that of the receiving estuarine waters, greatly influences the initial dilution. The most efficient initial dilution results when the densities of the contaminant and of the receiving estuarine waters are equal. A contaminant with a lower density than the surface waters of the estuary is initially confined to a thin surface lens, and the initial dilution as well as the subsequent diffusion is low. Likewise, poor dilution and slow diffusion result from the introduction of a contaminant with

a much higher density than the bottom water, since the waters are initially confined to a thin bottom layer.

Maximum initial dilution of wastes which are less dense than the receiving estuarine waters is achieved when the wastes are introduced near the bottom, since considerable entrainment of surrounding water occurs as the plume of contaminant ascends to the surface. Likewise, wastes that are denser than the receiving estuarine waters are subject to maximum initial dilution if introduced near the surface.

If a contaminant with an initial density equal to or greater than the receiving waters is introduced into the deep layers of a salt-wedge estuary, the contaminated volume disperses only very slowly within the wedge, since the currents there are weak. The waste is carried by the slow upstream-directed flow to the very tip of the wedge. The process of upward exchange of water from the wedge slowly transfers contaminated water to the seaward-flowing upper layer. Once in the upper layer, the waste is carried seaward with the surface water layers and transported out of the estuary. The time required to flush the wastes from the estuary in this case is relatively long, since the rate of transfer of the waste from the salt wedge to the surface layers is slow.

A contaminant introduced into the surface layers of a salt-wedge estuary, with an initial density equal to or less than the receiving waters, is rapidly flushed from the estuary with the seaward-flowing surface waters. None of the contaminated water enters the salt wedge, since exchange across the upper interface of the wedge is essentially one-way, directed from the wedge to the fresher surface layers.

A contaminant initially introduced into the bottom layers in a partially mixed coastal plain estuary, in addition to participating in the oscillatory movement of the tidal currents, is carried in the net motion toward the head of the estuary. At the same time, turbulent mixing leads to horizontal dispersion in both the longitudinal and lateral directions, and to vertical dispersion into the surface layers. The wastes which become mixed with the surface layers are carried in the net flow toward the mouth. Seaward from the point of introduction, the contaminant being carried toward the ocean in the surface layers is partially mixed downward into the deeper layers, and reintroduced into layers moving toward the head of the estuary.

A contaminant introduced into the surface layers is initially carried in the net flow toward the mouth of the estuary. Turbulent mixing leads to both horizontal and vertical dispersion, and the wastes are thus also added to the deeper layers having a net flow directed towards the head of the estuary.

In the region of the estuary headward from the point of introduction, the concentrations of the contaminant will always be greater in the deeper layers than in the surface layers, while seaward from the point of introduction the converse will be true. These conditions prevail regardless of whether the wastes are initially introduced into the surface layers or into the deeper layers.

The contaminant is ultimately flushed from the estuary in the seaward-directed flow of the surface layers.

Estuarine environments. Estuarine ecological environments are complex and highly variable compared to other marine environments. They are richly productive. A number of commercially important marine forms are indigenous to the estuary, and this environment serves as a spawning or nursery area for other prominent species.

River inflow and land drainage provide the primary source of nutrients required for the production of plant life in the estuary. Relative shallowness coupled with tidal mixing provides for the return of nutrients from decayed organic matter on the bottom to the productive surface layers.

Many estuaries are quite turbid as a result of suspended silt carried into the environment with the river inflow, as well as from shore erosion. The depth to which the solar energy necessary for photosynthesis can penetrate is thus limited to a relatively thin surface layer. However, a producing phytoplankton population may be maintained over a much thicker layer as a result of the stirring induced by the tidal motion and by the wind. This stirring keeps the phytoplankton in the layers of adequate light intensity for sufficient time to produce organic matter in excess of utilization. For a more complete treatment of the ecology of estuarine environments from a biological viewpoint, see MARINE ECOSYSTEM.

The close relationship between the circulation pattern and certain faunal distributions should be pointed out. The net headward flow in the deeper layers of a partially mixed coastal plain estuary is important in transporting certain larval forms up the estuary. Thus fingerling fish, notably the croaker *Micropogon undulatus*, which are spawned in the coastal waters off the mouth of coastal plain estuaries along the eastern coast of the United States, are carried into the estuary, where a suitable nursery environment exists. The flow in the bottom layers is also important in transporting oyster larvae from brood stock areas located in the lower reaches of some estuaries to oyster seed beds further up the estuary. [D.W.P.]

Estuarine sediments. The inorganic sediments of estuaries are derived from inflowing rivers, bordering sea cliffs, the sea floor outside the estuary, and the reworked deposits of tidal flats and marshes along the shores. Regardless of the source, much reworking of sediment occurs within estuaries and lagoons. Erosion, too, is evidenced by the migration of tidal channels and the muddy color of the water when no river inflow is taking place. Some estuaries have entrances narrow enough so that tidal currents scour the bottom locally, leaving rocky or gravelly bottoms. The prevailing condi-

tion, however, must be one of deposition and the average rate of deposition is greater than that of the open sea.

Distribution of grain sizes. The coarsest sediment in most estuaries is on the barrier or baymouth bar, and consists of sand and cobbles. Generally, this material is too coarse to have been transported across the tidal flats, but is derived from erosion of a sea cliff, then transported and deposited by longshore currents and waves. The excellent sorting and absence of much silt and clay may result from the turbulence of the waves. *See* SHORE PROCESSES.

The flat portions of the floors of estuaries that are deeper than about 18 ft are usually covered by sediment which becomes progressively finer with depth of water. A smooth concentric pattern of sediments may occur, ranging from sand along the shore to fine mud at depth. Such a distribution occurs only where the bottom is relatively flat and wave and current conditions are mild. In estuaries where the deeper areas are extremely irregular, mud occurs only in depressions and coarse sediments characterize the shallower bottoms.

The sediment distribution in shallow areas, mostly the tidal flats, is more complex but usually follows a systematic pattern. Most of the flow of water is confined to well-defined channels which slowly migrate over the tidal flat (as shown by remapping at intervals of several years). The velocity of the water is such that the finer grains are swept out, leaving the coarse sediment in the channel. The areas between the channels consist of poorly sorted mud which becomes finer with distance from the tidal channels. Probably most, but not all, of the reworking of sediment in estuaries takes place on the shallow flats where the ebbing and flooding currents erode and redeposit the sediment (Fig. 3).

Organic constituents. The sediments contain the remains of all phyla of animals and much plant debris. Even though the remains become scattered by scavenging, decomposition, and diagenesis, the organisms still have enriched the sediments in organic matter, calcium carbonate, silica, nutrients, and other constituents.

Sediments in estuaries located in areas where precipitation exceeds evaporation have organic nitrogen contents from 0.2 to 0.6%, and sediments in hypersaline areas, below 0.2%. The percentage used to differentiate between these areas, or 0.2% organic nitrogen, corresponds to about 1.7% organic carbon, or 2.9% total organic matter. Phosphorus is also abundant in sediments of normal environments, ranging from 0.1 to 0.4%.

Calcium carbonate is variable because of the presence or absence of shells and because of solution induced by acidic conditions. In coastal bays in temperate and arctic regions, calcium carbonate ranges between 0 and 6%, whereas in bays of tropical regions it is 10–47%.

Manner of deposition. In the tidal section of a fresh-water river a transition takes place, and the distribution of sediments may be quite variable and confused. When the estuary proper is reached there is some admixture of sea salts, and where the net upstream flow in lower layers occurs, there is a distinct change in sediment distribution. Finer sediments tend to be deposited in the channel (the reverse of conditions commonly found in river channels). In most streams, the bulk of suspended material probably is silt which is deposited directly out of suspension. Clay sizes, however, may be deposited through flocculation. The clays then fall to the deeper floors of the estuaries. Sediment may also travel down rivers at or near the surface in large floating floccules containing organic debris. When these settle to the bottom or are stranded by lowering water level, they are held by capillary action. Near the mouth of the estuary coarser sediments are again found in the channel as a result of wave action and because much of the silt load has already been deposited in the channel further upstream. *See* MARINE MARSH; MARINE SEDIMENTS.

[R.E.ST.]



Fig 3. Aerial photograph of tidal flats showing areas of pans, marshes, and vegetation between channels. Scott Head Island, England. *(Photograph by J. K. St. Joseph, Crown copyright reserved)*

Bibliography: A. B. Arons and H. Stommel, A mixing-length theory of tidal flushing, *Trans. Am. Geophys. Union,* 32(3):419–421, 1951; K. O. Emery and R. E. Stevenson, Estuaries and lagoons, *Geol. Soc. Am. Mem. 67,* 1:673–750, 1957; H. E. Landsberg (ed.), *Advances in Geophysics,* vol. 1, 1952; D. W. Pritchard, Estuarine circulation patterns, *ASCE Proc.,* 81(717), 1955; R. E. Stevenson and K. O. Emery, *Marshlands at Newport Bay, Calif.,* Allan Hancock Foundation Publs. Occas. Pap. 20, 1958.

Ethane

A member of the alkane or paraffin series of hydrocarbons, formula CH_3CH_3. It occurs in natural gas, but in much smaller quantities (5–20%) than methane, the principal component (50–90%). It

is a colorless, odorless, normally gaseous hydrocarbon having a freezing point of $-183.3°C$ and a boiling point of $-88.6°C$.

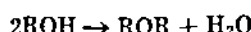
Ethane undergoes thermal reactions more readily than does methane. It begins to undergo dehydrogenation to ethylene and hydrogen at about 485°C; commercial operations are usually conducted at 800°C. As the temperature is raised, the reaction is accompanied by the formation of carbon, methane, acetylene, butadiene, and aromatic hydrocarbons. Pyrolysis of the ethane and propane portion of natural gas is used as an industrial method for production of ethylene. *See* ALKANE; CRACKING. [L.S.]

Ether

One of a class of organic compounds characterized by the structural feature of an oxygen atom linking two hydrocarbon groups, R—O—R′. They are used widely as solvents, both in chemical manufacture and in the research laboratory. The most important ether is ethyl ether, $C_2H_5OC_2H_5$.

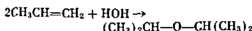
The hydrocarbon radicals, R and R′, may be identical (simple ether) or different (mixed ether). They may be aromatic or aliphatic, and the names of the ethers correspond to the hydrocarbon groups present. Thus, $CH_3—O—CH_3$ is methyl ether, rarely dimethyl ether, and $C_6H_5—O—CH_3$ is phenyl methyl ether.

Manufacture and preparation. Simple ethers may be considered to be the anhydrides of alcohols and are manufactured from alcohols by catalytic dehydration,



$$2ROH \rightarrow ROR + H_2O$$

or from olefins by controlled catalytic hydration,



$$2CH_3CH{=}CH_2 + HOH \rightarrow (CH_3)_2CH—O—CH(CH_3)_2$$

Mixed ethers of definite structure may be prepared by the Williamson synthesis,

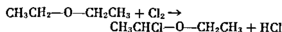


$$C_3H_7ONa + C_2H_5Br \rightarrow C_3H_7OC_2H_5 + NaBr$$

This synthesis was of considerable significance historically because a knowledge of the structure of ethers was important in developing the radical theory, a stepping stone to the present extensive knowledge concerning the arrangements of atoms in the molecules of organic compounds. A closely related reaction is that which takes place between cellulose, alkali, and ethyl chloride to yield an important plastic, the polyethyl ether of cellulose known as ethyl cellulose.

Properties. Ethers are less soluble in water than the corresponding alcohols, but are miscible with most organic solvents. Low-molecular-weight ethers have a lower boiling point than the corresponding alcohols, but for those ethers containing radicals larger than butyl, the reverse is true. The boiling points approximate those of hydrocarbons of the same molecular weight and geometry, indicating that association of ether molecules in the liquid state is negligible. Inertness at moderate temperatures, an outstanding chemical characteristic of the saturated alkyl ethers, leads to their wide use as reaction media. The organic magnesium compounds known as the Grignard reagents, RMgX, perhaps the most-used reagents in organic synthesis, are almost always prepared in ether solutions, and suspensions of alkali metals in ethers are often employed. At higher temperatures, however, ethers are split by the alkali metals and by the halide salts of metalloids.

Ethers may also be split by hydrogen halides. Hydrogen iodide (HI), for example, often reacts at room temperature to form an alcohol and an alkyl iodide. Ethers react with chlorine and bromine considerably more readily than do the corresponding hydrocarbons. The initial reaction involves the formation of hydrogen halide and the substitution of a halogen atom for one of the hydrogens on a carbon adjacent (alpha) to oxygen:



$$CH_3CH_2—O—CH_2CH_3 + Cl_2 \rightarrow CH_3CHCl—O—CH_2CH_3 + HCl$$

Because such α-halogens are reactive, the halogenated ethers are convenient intermediates for synthesis. Halogenated ethers are known in which the halogen is on a carbon other than that adjacent to the oxygen, but they are relatively inert.

On standing, ethers react with the oxygen of the air to form peroxides Before distillation, it is essential that any considerable accumulation of the latter be destroyed by alkaline hydrolysis, or by treatment with a reducing agent, such as ferrous hydroxide. On concentration and heating, ether peroxides detonate with dangerous violence. Distillation of ethers or ether solutions, particularly of those that have stood for some time, without taking precautions to test for and destroy peroxides, often leads to serious accidents. Some ethers form saltlike addition compounds with Lewis acids, the halogens, or picric acid, which are theoretically related to, but usually much less stable than, the corresponding derivatives of amines. This property permits the separation of ethers from inert hydrocarbons by extraction with concentrated sulfuric acid.

Identification of ethers is difficult. Often the more reactive components of a mixture are removed by chemical reagents, and the residual ethers identified by a combination of their failure to react and specific physical properties. The inertness of ethers is utilized in the syntheses of complicated organic molecules, an objectionably reactive alcohol group being protected by converting it to an unreactive ether. Hydrogen iodide may be used to regenerate the alcohol from the ether when the need for protection is passed

Not all ethers are inert. Unsaturated ethers undergo the reactions usually associated with the double bond Vinyl ethers, in which the double bond is adjacent to the oxygen, are readily polymerized or copolymerized with such monomers as vinyl acetate to yield useful polymers. Vinyl ethers also react in the presence of acid catalysts with

compounds that possess active hydrogens. Thus, with alcohols, they form acetals. Ethylene oxide is representative of another ether type which is quite reactive. *See* HETEROCYCLIC COMPOUNDS.

Ethyl ether. The best known of the ethers is ethyl ether, sometimes called diethyl ether or simply ether, $CH_3CH_2OCH_2CH_3$. It is used in industry as a solvent and in medicine as an anesthetic.

The older process of manufacture involved heating ethyl (grain) alcohol to moderate temperatures with catalytic quantities of sulfuric acid (H_2SO_4). Both ethyl ether and ethyl alcohol are now manufactured by the controlled catalytic hydration of ethylene, a by-product from the production of gasoline. The ethylene may be absorbed under pressure in H_2SO_4 to form sulfates, which are hydrolyzed to give mainly alcohol and some ether. Solid acid catalysts and a flow process are now generally used instead of the sulfuric acid method, and the proportion of alcohol to ether is controlled by variation of temperature and reactant concentrations. Thus, manufacturers may shift production to make predominantly either alcohol or ether. Obviously, the price is closely related to that of alcohol.

The anesthetic properties were first noticed by Paracelsus (1490–1541) and were rediscovered by Michael Faraday in 1818, but it was not until 1846 that its potential as a surgical anesthetic was demonstrated by W. T. G. Morton. The ethyl ether intended for anesthetic use differs from the ordinary variety in that possibly injurious impurities are removed. Peroxides are particularly harmful, and storage conditions must inhibit their formation.

When ether is used as a solvent, its high volatility can cause loss. However this volatility is advantageous in that ether can be readily removed from the concentrated or crystallized product. The toxicity to humans is low and recovery from overexposure is rapid and complete. It readily forms explosive mixtures with air, and on standing in containers which have been opened, it forms dangerous peroxides. Whereas ethyl ether is usually prepared from ethylene by partial hydration, the reverse process is used to supply ethylene to processing plants at a distance from the source. The ether is simply passed over alumina at 650°C to regenerate the ethylene and water almost quantitatively. Its freezing point is −117.4°C; boiling point, 34.6°C; density, 0.7146; refractive index, 1.35424. The solubility of ether in water is 6.18%; of water in ether, 1.2%.

Isopropyl ether. This is the cheapest industrial ether, and it is widely used to extract substances from water solutions, and to remove wax by crystallization from lubricating oil. It is a solvent for fats, waxes, and resins. The formula is $(CH_3)_2CHOCH(CH_3)_2$, and it is sometimes called diisopropyl ether.

It is manufactured from propylene by controlled catalytic hydration in a process in which isopropyl alcohol is usually the main product. Because the supply from this source exceeds the demand for chemical manufacturing uses, the excess is blended with gasoline. Its boiling point is 68.5°C; melting point, −86°C; refractive index, 1.3681; specific gravity, 0.724; water solubility, 1% at room temperature. *See* DIOXANE; ETHYLENE OXIDE; PROPYLENE. [R.K.S.]

Ether hypothesis

James Clerk Maxwell and his contemporaries in the nineteenth century found it inconceivable that a wave motion should propagate in empty space. They therefore postulated a medium, which they called the ether, that filled all space and transmitted electromagnetic vibrations.

During the last half of the nineteenth century, dozens of models were tried, but all broke down at some point. Direct experimental attempts to establish the existence of an absolute ether frame of reference, in which Maxwell's equations hold and light has the velocity c, have failed. The best known of these is the Michelson-Morley experiment, in which an attempt was made to measure the velocity of the earth relative to the ether. *See* LICHT.

Every hypothesis such as ether drag, Lorentz contraction, and so on, invented to reconcile some experiment with the ether concept, has been disproved by some other experiment. At present, there is no evidence whatever that the ether exists. *See* MAXWELL'S EQUATIONS. [W.R.SM.]

Bibliography: W. K. H. Panofsky and M. Phillips, *Classical Electricity and Magnetism*, 1955; E. Whittaker, *A History of the Theories of Aether and Electricity*, vol. 1, 1951.

Ethoxide

A derivative of ethanol (CH_3CH_2OH, ethyl alcohol). Ethoxides are formed by replacing the hydroxyl hydrogen of ethanol by a metal or an organic radical. The metallic derivatives, also known as ethylates, are made from ethanol and an active metal. Sodium and aluminum ethylates are hygroscopic white solids used as catalysts and in organic syntheses. The simpler organic derivatives of ethoxides are commonly known as ethers, and are prepared by the Erlenmeyer or Williamson syntheses. Although ethers are frequently regarded as derivatives of ethyl alcohol, they more closely resemble the hydrocarbons in activity. These are used as solvents and anesthetics. The less simple derivatives are widely distributed in nature as in vanillin, saponins, anthocyanins, flavones, and pigments. *See* ETHER; ETHYL ALCOHOL. [E.H.H.]

Ethyl acetate

The ethyl ester of acetic acid, formula



$$CH_3COOC_2H_5$$

melting point −83°C, boiling point 77°C. Ethyl acetate is a colorless, neutral, volatile, flammable liquid with a characteristic pleasant fruity odor; its solubility in water is 1 part in 10 by volume. Because it readily absorbs water and is easily hydro-

lyzed, it must be kept tightly sealed and cool in storage.

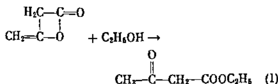
Ethyl acetate is manufactured in large quantities by the reaction of ethyl alcohol with acetic acid, acetic anhydride, or ketene; it can also be made via the Tischenko reaction from acetaldehyde.

Large amounts are used as a solvent in varnishes, nitrocellulose lacquers, and airplane dopes; other uses include artificial flavors, perfumes, acetate rayon manufacture, smokeless powder, artificial leather, and photographic film. *See* ACID ANHYDRIDE; ESTER; ESTERIFICATION; SOLVENT. [E.B.R.]

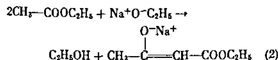
Ethyl acetoacetate

The ethyl ester of the thermally unstable acetoacetic acid (hence often called acetoacetic ester) is a colorless, stable oil with the molecular formula $C_6H_{10}O_3$, melting point −45°C, boiling point 180°C, at 755 mm, and specific gravity 1.025_4^{20}. First prepared by A. Geuther in 1863, ethyl acetoacetate has been much studied; as a result, it has proved to be a versatile chemical compound, and is used as an intermediate in the synthesis of a variety of important substances.

Preparation. Ethyl acetoacetate is commonly synthesized by two important methods:
From diketene and ethyl alcohol

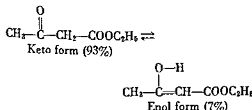


From ethyl acetate, via the Claisen condensation



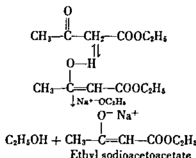
In the latter, ethyl sodioacetoacetate is the product. Acidification of this liberates the free ester, which is purified by distillation.

The ester is difunctional and exists in both of two tautomeric forms, the one a ketone and the other an enol, dynamically related to each other by the following equilibrium:

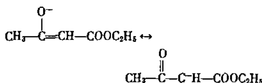


in which the relative proportions of keto and enol forms present at room temperature are approximately as indicated. Both forms can be isolated, although neither can be kept free from the other indefinitely. L. Knorr (1911) found that the keto form separated from alcoholic or petroleum ether solutions at −78°C; when freshly tested, it failed to give a color with ferric chloride solution, and was inert to bromine. The enol form was obtained by passing dry hydrogen chloride into a suspension of the sodium salt of acetoacetic ester at −78°C; a glassy solid formed that reacted instantly with bromine and with ferric chloride. On standing, however, each form rearranged into the other, the rearrangement being catalyzed by minute traces of either acids or bases. A second method for isolating the enol in pure form (K. H. Meyer, 1920) consists of slow distillation of the ester from aseptically clean quartz apparatus. Studies have shown that the keto form rearranges slowly into the enol, and that the enol form, as opposed to the keto, reacts quickly and quantitatively with bromine; the proportion of enol in equilibrium with ketone can therefore be determined directly by a quick titration of the mixture (at 0°C) with standard bromine solution (Meyer enol test). *See* TAUTOMERISM.

When exposed to the organic base, sodium ethoxide, acetoacetic ester behaves as an acid (through the enol form), furnishing ethyl alcohol and the salt, ethyl sodioacetoacetate.

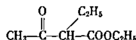


The anion of this salt is capable of resonance involving both the oxygen and the α-carbon atoms.



and the actual structure of the anion is a hybrid of the above two ionic forms. *See* RESONANCE (MOLECULAR STRUCTURE).

Ethyl sodioacetoacetate reactions. Of great synthetic value is the fact that when the sodio-salt is refluxed in alcohol with a primary alkyl bromide or iodide, it undergoes alkylation on the α-carbon atom. Thus, ethyl ethylacetoacetate,



forms by alkylation with ethyl iodide. This ester still contains an active α-hydrogen atom; it is therefore capable of enolization and of one further alkylation under the above conditions, to give finally the nonenolic, dialkylated acetoacetic ester,

compounds that possess active hydrogens. Thus, with alcohols, they form acetals. Ethylene oxide is representative of another ether type which is quite reactive. *See* HETEROCYCLIC COMPOUNDS.

Ethyl ether. The best known of the ethers is ethyl ether, sometimes called diethyl ether or simply ether, $CH_3CH_2OCH_2CH_3$. It is used in industry as a solvent and in medicine as an anesthetic.

The older process of manufacture involved heating ethyl (grain) alcohol to moderate temperatures with catalytic quantities of sulfuric acid (H_2SO_4). Both ethyl ether and ethyl alcohol are now manufactured by the controlled catalytic hydration of ethylene, a by-product from the production of gasoline. The ethylene may be absorbed under pressure in H_2SO_4 to form sulfates, which are hydrolyzed to give mainly alcohol and some ether. Solid acid catalysts and a flow process are now generally used instead of the sulfuric acid method, and the proportion of alcohol to ether is controlled by variation of temperature and reactant concentrations. Thus, manufacturers may shift production to make predominantly either alcohol or ether. Obviously, the price is closely related to that of alcohol.

The anesthetic properties were first noticed by Paracelsus (1490–1541) and were rediscovered by Michael Faraday in 1818, but it was not until 1846 that its potential as a surgical anesthetic was demonstrated by W. T. G. Morton. The ethyl ether intended for anesthetic use differs from the ordinary variety in that possibly injurious impurities are removed. Peroxides are particularly harmful, and storage conditions must inhibit their formation.

When ether is used as a solvent, its high volatility can cause loss. However this volatility is advantageous in that ether can be readily removed from the concentrated or crystallized product. The toxicity to humans is low and recovery from overexposure is rapid and complete. It readily forms explosive mixtures with air, and on standing in containers which have been opened, it forms dangerous peroxides. Whereas ethyl ether is usually prepared from ethylene by partial hydration, the reverse process is used to supply ethylene to processing plants at a distance from the source. The ether is simply passed over alumina at 650°C to regenerate the ethylene and water almost quantitatively. Its freezing point is −117.4°C; boiling point, 34.6°C; density, 0.7146; refractive index, 1.35424. The solubility of ether in water is 6.18%; of water in ether, 1.2%.

Isopropyl ether. This is the cheapest industrial ether, and it is widely used to extract substances from water solutions, and to remove wax by crystallization from lubricating oil. It is a solvent for fats, waxes, and resins. The formula is $(CH_3)_2CHOCH(CH_3)_2$, and it is sometimes called diisopropyl ether.

It is manufactured from propylene by controlled catalytic hydration in a process in which isopropyl alcohol is usually the main product. Because the supply from this source exceeds the demand for chemical manufacturing uses, the excess is blended with gasoline. Its boiling point is 68.5°C; melting point, −86°C; refractive index, 1.3681; specific gravity, 0.724; water solubility, 1% at room temperature. *See* DIOXANE; ETHYLENE OXIDE; PROPYLENE. [R.K.S.]

Ether hypothesis

James Clerk Maxwell and his contemporaries in the nineteenth century found it inconceivable that a wave motion should propagate in empty space. They therefore postulated a medium, which they called the ether, that filled all space and transmitted electromagnetic vibrations.

During the last half of the nineteenth century, dozens of models were tried, but all broke down at some point. Direct experimental attempts to establish the existence of an absolute ether frame of reference, in which Maxwell's equations hold and light has the velocity *c*, have failed. The best known of these is the Michelson-Morley experiment, in which an attempt was made to measure the velocity of the earth relative to the ether. *See* LIGHT.

Every hypothesis such as ether drag, Lorentz contraction, and so on, invented to reconcile some experiment with the ether concept, has been disproved by some other experiment. At present, there is no evidence whatever that the ether exists. *See* MAXWELL'S EQUATIONS. [W.R.SM.]

Bibliography: W. K. H. Panofsky and M. Phillips, *Classical Electricity and Magnetism*, 1955; E. Whittaker, *A History of the Theories of Aether and Electricity*, vol. 1, 1951.

Ethoxide

A derivative of ethanol (CH_3CH_2OH, ethyl alcohol). Ethoxides are formed by replacing the hydroxyl hydrogen of ethanol by a metal or an organic radical. The metallic derivatives, also known as ethylates, are made from ethanol and an active metal. Sodium and aluminum ethylates are hygroscopic white solids used as catalysts and in organic syntheses. The simpler organic derivatives of ethoxides are commonly known as ethers, and are prepared by the Erlenmeyer or Williamson syntheses. Although ethers are frequently regarded as derivatives of ethyl alcohol, they more closely resemble the hydrocarbons in activity. These are used as solvents and anesthetics. The less simple derivatives are widely distributed in nature as in vanillin, saponins, anthocyanins, flavones, and pigments. *See* ETHER; ETHYL ALCOHOL. [E.H.H.]

Ethyl acetate

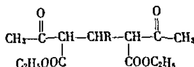
The ethyl ester of acetic acid, formula



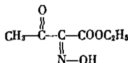
$$CH_3COOC_2H_5$$

melting point −83°C, boiling point 77°C. Ethyl acetate is a colorless, neutral, volatile, flammable liquid with a characteristic pleasant fruity odor; its solubility in water is 1 part in 10 by volume. Because it readily absorbs water and is easily hydro-

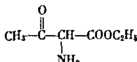
nally the diketo diester,



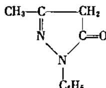
Ethyl acetoacetate undergoes nitrosation on the α-carbon when treated with nitrous acid, to give isonitrosoacetoacetic ester,



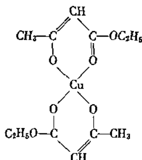
Reduction of the latter furnishes α-aminoacetoacetic ester



an important intermediate in the Knorr synthesis of pyrroles (*see* PYRROLE). With hydrazine, condensation and cyclization occur to form a dihydropyrazolone. Thus, acetoacetic ester and phenylhydrazine react to form the heterocyclic compound, 1-phenyl-3-methyl-5-pyrazolone,



Certain divalent metals, for example, copper, are capable of forming stable, crystalline, covalent chelate compounds with acetoacetic ester; these possess the following type of structure:



Acetoacetic ester shows many of the normal carbonyl reactions expected for a ketone, including cyanohydrin formation and sodium bisulfite addition. See CHELATION; CONDENSATION REACTION; ESTER; HYDROLYSIS; ORGANIC CHEMICAL SYNTHESIS.

[E.B.R.]

Ethyl alcohol

Probably the best known of the alcohols. It is also known as alcohol, ethanol, grain alcohol, industrial alcohol, fermentation alcohol, cologne spirits, ethyl hydroxide, and methylcarbinol. Pure ethyl alcohol is a colorless, limpid, volatile liquid which is flammable, is toxic, and has a pungent taste. It boils at 78.4°C and melts at −112.3°C, has a specific gravity of 0.7851 at 20°C, and is soluble in water and most organic liquids. It is one of the most important industrial organic chemicals. Billions of pounds of it are produced annually. Ethyl alcohol is produced by chemical synthesis and by fermentation or biosynthetic processes.

Uses. Ethyl alcohol is used as a solvent, extractant, antifreeze, and intermediate in the synthesis of innumerable organic chemicals. It is also, of course, an essential ingredient of alcoholic beverages. Its importance as an industrial organic chemical may be judged from the fact that over 1,270,000,000 lb of synthetic ethyl alcohol alone was produced in the United States in 1956.

Various grades of ethyl alcohol are produced, depending on their intended use. U.S. Pharmaceutical (U.S.P. XV) grade is the water azeotrope of ethyl alcohol and is 95% ethyl alcohol by volume. National Formulary (N.F. X) grade is 99+% ethyl alcohol by weight; it is also called absolute, or anhydrous, alcohol. This grade is generally prepared by azeotropic dehydration with benzene and therefore usually contains about 0.5% benzene. Denatured alcohol contains a small amount of a malodorous or obnoxious material to prevent the use of this nontaxed grade of ethyl alcohol for beverage purposes Ethyl alcohol which is employed in the manufacture of beverages, medicine, and flavoring is taxed and all ethyl alcohol production is closely supervised by the government.

The concentration of ethyl alcohol is often expressed as proof, which is simply twice the volume per cent of ethyl alcohol. Thus, a 100-proof whiskey contains 50% by volume of ethyl alcohol.

The major use of ethyl alcohol is as a starting material for various organic syntheses, Bimolecular dehydration of ethyl alcohol gives diethyl ether, which is employed as a solvent, extractant, and anesthetic. Dehydrogenation of ethyl alcohol yields acetaldehyde which is the precursor of a vast number of organic chemicals such as acetic acid, acetic anhydride, chloral, butanol, crotonaldehyde, and ethylhexanol. Reaction with carboxylic acids or anhydrides yields esters which are useful in many applications. The hydroxyl group of ethyl alcohol may be replaced by halogen to give the ethyl halides. Treatment with sulfuric acid gives ethyl hydrogen sulfate and diethyl sulfate, a useful ethylating agent. Reaction of ethyl alcohol with aldehydes gives the respective diethyl acetals, and reaction with acetylene produces the acetals, as well as ethyl vinyl ether. Treatment of ethyl alcohol with ammonia produces acetonitrile, which may be reduced to ethylamine. These and other ethyl-alcohol-derived chemicals are used in dyes, drugs, synthetic rubber, solvents, extractants, detergents, plasticizers, lubricants, surface coatings, adhesives, moldings, cosmetics, explosives, pesticides, and [illegible]thetic fiber resins.

Flow sheet of industrial ethyl alcohol manufacture. (From S. C. Prescott and C. G. Dunn, Industrial Microbiology, 2d ed., McGraw-Hill, 1949)

Synthetic production. Ethyl alcohol is produced commercially by several synthetic processes.

Catalytic hydration of ethylene. Ethylene, produced by petroleum cracking and from natural gas, is treated with water at high temperatures in the presence of acidic catalysts to produce ethyl alcohol.

Sulfuric acid hydration of ethylene. Ethylene is treated with concentrated sulfuric acid to produce ethyl hydrogen sulfate and diethylsulfate which are then hydrolyzed to ethyl alcohol and dilute sulfuric acid.

Fischer-Tropsch process. Ethyl alcohol is a major by-product in the synthesis of methanol by the reaction of carbon monoxide and hydrogen over iron catalysts. [J.W.L.]

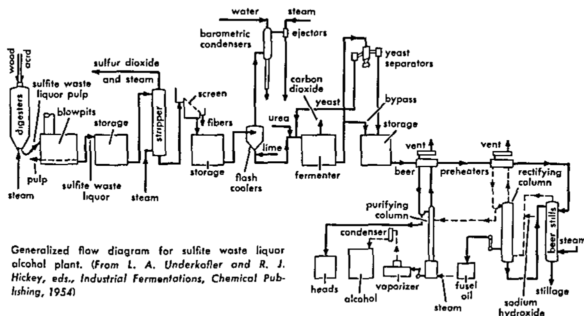
Industrial fermentation. The chemical produced in largest volume by industrial fermentation is ethyl alcohol (ethanol, C_2H_5OH). Ethyl alcohol is produced only from hexose sugars by yeast according to the over-all reaction

$$\underset{\text{Hexose}}{C_6H_{12}O_6} \rightarrow \underset{\text{Ethanol}}{2C_2H_5OH} + \underset{\text{Carbon dioxide}}{2CO_2}$$

The highest yield of alcohol obtainable is about 96% of the theoretical. The theoretical yield is about 51% by weight, based on the sugar fermented. Disaccharides, such as sucrose and maltose, are first split by hydrolytic enzymes of the yeast cell into monosaccharides, the hexose sugars. Polysaccharides, such as starch and cellulose, cannot be fermented directly by yeast, but must be hydrolyzed to simple sugars. This is usually done by the use of enzymes and sometimes by heating with a mineral acid. The most common substrates for the production of industrial alcohol are molasses, grain, sulfite waste liquor, and wood waste. The fermentation is carried out for the most part in the absence of air. This decreases the yield of cells and the amount of carbon dioxide produced from respiration, and it increases the yield of alcohol. The alcohol is finally purified and concentrated by distillation.

Molasses alcohol. This is a relatively simple process, since about 90% of the sugars present in blackstrap molasses is in fermentable form, as sucrose and invert sugar. In so-called high-test or invert molasses, about 95% of the sugars is fermentable by yeast. Molasses is diluted with water to a sugar content of 14–18% and pumped directly into the fermenter. During filling 2–4% of an actively fermenting yeast mash is also added. The filling may require about 8 hours, during which time the yeast multiplies actively. The pH is adjusted with sulfuric, lactic, or hydrochloric acid to 4–5. Sometimes a little ammonium sulfate is added to supply nitrogen for yeast growth, but generally molasses contains sufficient minerals to allow ample yeast growth.

Fermentation produces 26 kilocalories of heat per mole of hexose fermented, and if fermenters



Generalized flow diagram for sulfite waste liquor alcohol plant. *(From L. A. Underkofler and R. J. Hickey, eds., Industrial Fermentations, Chemical Publishing, 1954)*

are therefore cooled to a temperature of 80–90°F. If this is not done, alcohol production is impaired. Fermentation is complete after 36–72 hours, depending on the type of molasses used. Usually 2.3–2.7 gal of blackstrap molasses are required to produce 1 gal of 190-proof alcohol. The principal by-products of the fermentation are carbon dioxide, produced in yields almost as high as alcohol, and fusel oil, the high-boiling fraction of the distillation. The evaporated and concentrated molasses residue can be used as stock feed. The nonconcentrated stillage residue can be used directly as a soil fertilizer. Continuous processes, rather than the batch method, are also used for industrial alcohol production.

Grain alcohol. Most distilleries use distiller's barley malt as a source of the enzyme amylase to convert the starch to fermentable sugars. After the grain is milled, the flour is cooked to hydrate and gelatinize the starch. Cooking is often done in pressure cookers, with a steam pressure of 100 psi. Both batch and continuous processes are used. After partial cooking, distiller's malt of high diastatic power is added so that the mixture reaches a temperature of about 145°F. The starch is liquefied rapidly and about 75–80% is converted to maltose and other fermentable sugars and the remainder to residual dextrins, called limit dextrins. In some countries, conversion of the starch is accomplished by the use of a diastatic mold, *Rhizopus delemar.* The mold is grown in the mash for 24 hours prior to inoculation with yeast. The mash is cooled from 145°F to 70–85°F by the use of heat exchangers. Yeast nutrients are supplied by adding 20–25% stillage to the volume of cooled mash. Fermentation and distillation take place essentially as described previously. The residual dextrins are slowly converted to maltose during the 2–3 day fermentation period.

Sulfite waste liquor. This by-product is also a suitable source of carbohydrates for industrial alcohol fermentation by yeast. Sulfite waste liquor from softwoods, or gymnosperms, is preferred over that from hardwood, or angiosperms, since the former has a much lower content of pentoses which are not fermentable by yeast. Since the content of fermentable sugars is only about 1–2%, the waste liquor is used without dilution. Preparation of the liquor for fermentation is similar to the process described for the production of fodder yeast from this source. Yields varying from 10–40 gal of 95% alcohol per ton of air-dried pulp can be obtained.

Wood waste as a source of alcohol. The polysaccharides of wood are converted to fermentable sugars by acid hydrolysis (*see* YEAST, INDUSTRIAL). Since use is made of the cellulose fraction of wood after hydrolysis to glucose, alcohol yields are much higher than those from sulfite waste liquor. Yields of 19–22 liters of alcohol per 100 kg of wood have been reported. *See* ALCOHOL; INDUSTRIAL MICROBIOLOGY. [H.J.P.; E.M.M.]

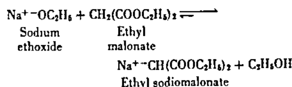
Bibliography: L. A. Underkofler and R. J. Hickey (eds.), *Industrial Fermentations,* vol. 1, 1954.

Ethyl malonate

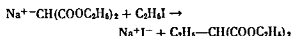
The diethyl ester of malonic acid, a dicarboxylic acid first obtained as an oxidation product of malic acid. The ester, $CH_2(COOC_2H_5)_2$, is liquid at room temperature, has specific gravity 1.055_4^{20}, freezing point −51.5°C, and boiling point 199.3°C. Of unusual synthetic value, the ester is manufactured commercially from acetic acid. The latter is chlorinated, transformed into sodium chloracetate ($ClCH_2COO^-Na^+$), and brought into reaction with sodium cyanide to form sodium cyanoacetate ($Na^{+-}OOCCH_2CN$). Sodium cyanoacetate may be either hydrolyzed with dilute sulfuric acid to give malonic acid, which is then esterified, or alc

ized directly to ethyl malonate using ethyl alcohol-sulfuric acid.

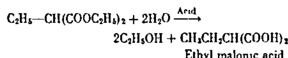
Alkylation reactions. The unusual value of malonic esters [$CH_2(COOR)_2$] lies in the fact that they are incipient acids, capable of largely neutralizing a strong organic base such as sodium ethoxide, according to the following:



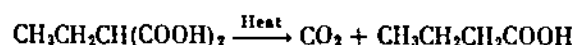
The sodiomalonate ion is capable of a variety of alkylation reactions, carried out usually by heating with the desired primary halide. Thus, with ethyl iodide



an ethylated malonic ester is produced, and this, in common with all substituted or unsubstituted malonic esters, can be hydrolyzed to the thermally unstable ethyl malonic acid

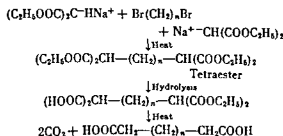


When ethyl malonic acid is heated above its melting point, it undergoes decarboxylation to furnish, in this case, *n*-butyric acid:

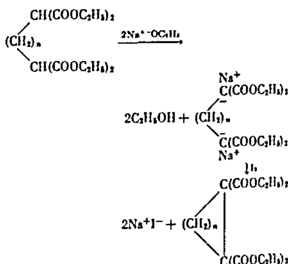


Analogously, any mono- or dialkylated malonic ester can be hydrolyzed and decarboxylated to give, respectively, mono- or disubstituted acetic acids; many long-chain and branched-chain acids have been prepared by this method. If it is desired to obtain the substituted malonic acid, the alkylated ester is first saponified with sodium or potassium hydroxide, and the resulting salt is carefully acidified with hydrochloric or sulfuric acids to liberate the organic acid

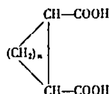
Malonic ester also is much used in the synthesis of long-chain $\alpha,\alpha,\omega,\omega$-tetracarboxylic esters, and from these by hydrolysis and decarboxylation, derived α,ω-dicarboxylic acids,



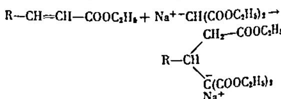
If in the tetraester, $n = 1, 2, 3,$ or 4, treatment with two equivalents of sodium ethoxide followed by addition of iodine causes cyclization. By this means, tetraesters of alicyclics containing three, four, five, or six carbon atoms can be made. Thus,



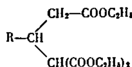
The tetraester can be hydrolyzed and decarboxylated (as described before), to yield the derived alicyclic dicarboxylic acid, of general formula



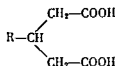
Ethyl sodiomalonate, with α,β-unsaturated esters, adds across the double bond (Michael reaction), according to the following:



from which, on neutralization with mineral acid, the triester

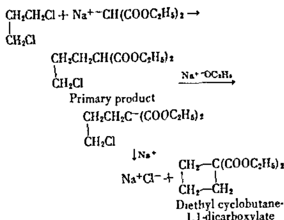


is obtained. Hydrolysis and decarboxylation of this triester furnishes substituted glutaric acids,



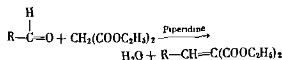
In the presence of an α,ω-dihalide (terminal dihalide), of two, three, four, or five carbons in length, the primary alkylation product, on treatment with additional sodium ethoxide, undergoes a

secondary alkylation-cyclization reaction, according to the following:



Through use of suitable dihalides, rings containing three, four, five, or six carbon atoms can be made.

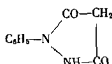
Condensation reactions. Condensations, with dehydration, involving esters of malonic acid and aldehydes, can be effected in the presence of piperidine or diethylamine. The products are α,β-unsaturated dicarboxylic esters



Malonic esters, when condensed with urea, give barbituric acids (*see* UREID); analogously, when condensed with thiourea, thiobarbituric acid results (*see* THIOUREA). Treatment with nitrous acid gives ethyl nitrosomalonate, $NOCH(COOC_2H_5)_2$, and reduction of this furnishes ethyl aminomalonate, $NH_2CH(COOC_2H_5)_2$. The latter is a valuable synthetic intermediate much used in the preparation of amino acids (*see* AMINO ACIDS). Malonic ester is also used in preparing methone (5,5-dimethyl-1,3-cyclohexanedione), a substance of unusual value in the characterization of aldehydes (*see* ALDEHYDE). With phenylhydrazine



and sodium ethoxide, malonic ester undergoes cyclization, furnishing phenyl diketopyrazolidine



The active methylene group ($-CH_2-$) of malonic ester is easily substituted by halogens (chlorine or bromine) to give corresponding halomalonic esters, or oxidized by nitrous gases (N_2O_4), to yield oxomalonic ester, $O{=}C(COOC_2H_5)_2$; such compounds are of theoretical and synthetic value. *See* CONDENSATION REACTION; ESTER; ORGANIC CHEMICAL SYNTHESIS. [E.B.R.]

Ethylene

A colorless gas, formula C_2H_4, boiling point −103.8°C, melting point −169.4°C. Ethylene is one of the major growth chemicals in the United States. Annual production is estimated to be over 3,000,000,000 lb. Breakdown on end use is ethyl alcohol 30%, ethylene oxide 25%, styrene 10%, ethyl chloride 10%, polyethylene 15%. Ethylene dichloride and other smaller uses account for the remainder. Most processes for the production of ethylene involve the cracking of either ethane or propane. High temperatures, short contact time in the cracker, and a partial pressure lowered by the use of diluents such as steam give maximum yield. When naphthas are used as feedstock, diolefins and aromatics are obtained along with ethylene.

Recovery of ethylene from the cracked gas is accomplished in most commercial installations by fractionation at low temperature using pressures in the order of 500–600 psig. Other systems employ solvent extraction, absorption, and low-temperature–low-pressure fractionation. Equipment for the latter process is made of alloy steels, although copper and aluminum have also been used.

The development of the Ziegler low-pressure–high-density polyethylene has given a tremendous impetus to increased production in this field. It is expected that a major outlet for ethylene in the future will be in the field of polymers and plastics. *See* ALKENE; POLYOLEFIN RESINS. [C.A.C.]

Ethylene chlorohydrin

A colorless, mobile liquid with characteristic ethereal odor, boiling point (760 mm Hg) 128.7°C, melting point, −62.6°C. It forms, with water, a constant-boiling mixture which boils at 97.8°C and contains 42.3% by weight of the chlorohydrin. Commercial production is in the order of 500,000,000 lb per year by simultaneous addition of chlorine and ethylene into a packed tower through which a countercurrent stream of water is circulated. A dilute solution (5–8%) of the chlorohydrin is continuously withdrawn and concentrated to the constant boiling mixture by distillation. Approximately 5–10% of the ethylene is converted to ethylene dichloride which is sold as a solvent and fumigant or converted to vinyl chloride.

Major use for the chlorohydrin is in the production of ethylene oxide which is made on a large scale for the manufacture of synthetic detergents, ethylene glycol for antifreeze and dynamite, polyester fibers, resins, and urethane foams. *See* ALKENE; ETHYLENE. [C.A.C.]

Ethylene glycol

A colorless, almost odorless liquid with a sweet taste, formula $HOCH_2CH_2OH$. It is relatively nonvolatile and viscous and is soluble in water and many organic liquids. Its freezing point is 9°F. Mixtures with water (58–80% glycol) freeze below −50°F. The exact eutectic composition is difficult to define.

It is the simplest and most widely used of the glycols. United States production of this chemical has increased from 2,000,000 lb in 1925 to over 1,000,000,000 lb in 1957.

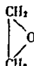
The largest use for ethylene glycol is as an automotive antifreeze, and for this purpose, inhibitors are generally added to prevent the formation of corrosive oxidation products. It is widely used in making explosives and is added to cellophane as a softening agent. It is an ingredient of hydraulic-brake and shock-absorber fluids and is the starting material for the chemical synthesis of drugs, plasticizers, and synthetic fibers.

Most of the glycol is manufactured by the hydration of ethylene oxide either at elevated temperatures in the absence of a catalyst or at lower temperatures in the presence of small amounts of strong acid catalysts. An alternate process involves hydrolysis of ethylene chlorohydrin. This is economical if a source of chlorine is at hand. It is also manufactured by hydrogenating methyl glycolate at high temperatures and at pressures above 400 atm. Ethylene glycol can be prepared in the laboratory by alkaline hydrolysis of ethylene dichloride, ethylene dibromide, or ethylene glycol diacetate. Other synthetic methods for making the glycol include the aluminum amalgam reduction of glycolaldehyde and the hydroxylation of ethylene with hydrogen peroxide or potassium permanganate.

Ethylene glycol takes part in many reactions common to monohydric alcohols, for example, etherification, esterification, condensation, and oxidation. Mono- or dialkyl ethers of ethylene glycol are formed by its reaction with sodium hydroxide and dialkyl sulfates or with sodium alcoholates and alkyl halides. Dioxane, a cyclic diether, is prepared by acid-catalyzed dehydration of ethylene glycol. Diethylene glycol, triethylene glycol, and higher polyethylene glycols are produced by the alkaline-catalyzed condensation of ethylene oxide with glycol. Most aldehydes and ketones condense with ethylene glycol to form cyclic acetals called dioxolanes, and with the dialdehyde glyoxal, O=HC—CH=O, it forms a bicyclic acetal, *p*-dioxano-(*b*)-*p*-dioxane. Vapor-phase dehydrogenation and dehydration of glycol over copper-chromium catalysts give hydroxymethyl dioxolane. Oxidation of glycol produces various products—oxalic acid, glycolic acid, formaldehyde, glycolic aldehyde, or glyoxal—depending upon the reaction conditions. Mono- and diesters of ethylene glycol are formed by its reaction with acids, acid anhydrides, or acid chlorides. With nitric acid, glycol forms the dinitrate, which is employed in conjunction with nitroglycerin to give explosives with low freezing points. Dibasic acids or anhydrides react with glycol to form polyesters; such acids as maleic, succinic, phthalic, and terephthalic have served to produce useful polymers of this type. *See* ANTIFREEZE MIXTURE; GLYCOL; POLYESTER RESIN; POLYETHYLENE GLYCOL; POLYHYDROXY ALCOHOL. [J.T.A.]

Ethylene oxide

The simplest cyclic ether or epoxide. It is also called epoxyethane. Some is used as a grain fumigant, but the major portion serves as a chemical



intermediate in the manufacture of ethylene glycol (the common permanent antifreeze), nonionic emulsifying agents, plastics, plasticizers, one type of synthetic rubber, and several important synthetic textiles. Another important use is as a gaseous sterilizing agent. For this purpose it is mixed with CO_2 (10% ethylene oxide, 90% CO_2) and put into an air-tight container with the material to be sterilized.

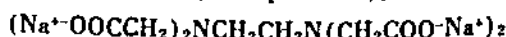
In one method of manufacture, ethylene is treated with hypochlorous acid to form ethylene chlorohydrin, which is converted to ethylene oxide by milk of lime. In a newer process, the ethylene is oxidized over a silver catalyst directly by a limited amount of air. Since ethylene is a cheap by-product of gasoline manufacture, and the oxidation process is surprisingly efficient, the potential supply is enormous.

The boiling point of ethylene oxide is 10.4°C; melting point, −112°C; refractive index, 1.3597; specific gravity, 0.87. It is soluble in water in all proportions. Unlike ordinary aliphatic ethers, it reacts readily in the presence of traces of acid or base with molecules containing an active hydrogen, such as water, hydrogen sulfide, ammonia, hydrogen cyanide, amines, and alcohols. The reaction with water to yield ethylene glycol, $HOCH_2CH_2OH$, is typical. It is used in classical organic syntheses with a Grignard reagent to lengthen a carbon chain by two units. *See* DIOXANE; ETHER; ETHYLENE; HETEROCYCLIC COMPOUNDS. [R.K.S.]

Bibliography: G. O. Curme and F. Johnston, *Glycols*, ACS Monograph 114, 1952.

Ethylenediaminetetraacetic acid

A substance, the sodium salt of which is very useful in forming complex compounds. The tetrasodium salt (EDTA, versene, or sequestrene),


$$(Na^{+-}OOCCH_2)_2NCH_2CH_2N(CH_2COO^-Na^+)_2$$

is a white powder, and is very soluble in water, in which it gives a basic reaction. Prepared either from ethylenediamine, formaldehyde, and sodium cyanide in basic solution, or from ethylenediamine and sodium chloroacetate, EDTA is a strong complexing or chelating agent, reacting with many metallic ions to form soluble, nonionic chelates. As such, it is widely used in analysis to retain alkaline earths and heavy metals in solution. The calcium disodium salt is often used in pharmaceuticals to

prevent calcium depletion of the body during therapy. *See* CHELATION; COMPLEX COMPOUNDS.

[E.B.R.]

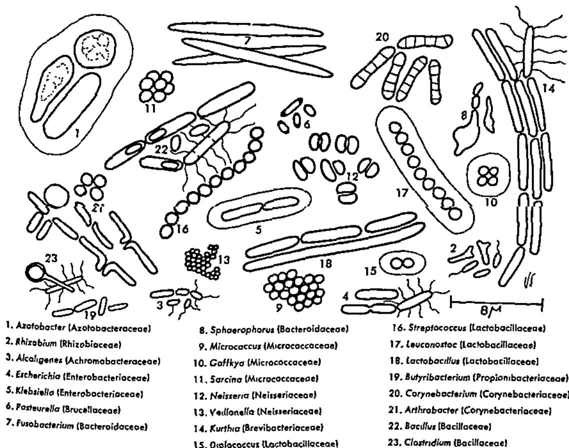
Eubacteriales

An order of bacteria of the class Schizomycetes, division Protophyta. *See* PROTOPHYTA.

Relationship with other orders. The name originally meant the true bacteria and suggests some sharp difference between bacteria of this order and those of other orders. When it was introduced by R. E. Buchanan in 1917, it was intended to separate all the common bacteria from those which (1) formed sheaths (Chlamydobacteriales), (2) branched (Actinomycetales), (3) deposited sulfur internally or produced bacteriochlorophyll (Thiobacteriales), (4) were flexible creeping cells which produced fruiting bodies (Myxobacterales), and (5) were slender flexuous spirals (Spirochaetales). *See* ACTINOMYCETALES; CHLAMYDOBACTERIALES; MYXOBACTERALES; SPIROCHAETALES.

The Spirochaetales, Myxobacterales, Actinomycetales, and Chlamydobacteriales have remained essentially the same. However, the species placed by Buchanan in the Thiobacteriales have been redistributed between two orders, the Pseudomonadales and Beggiatoales. Those species of the original order Eubacteriales, which then included all the single-celled organisms (excepting those with photosynthetic pigments or internal sulfur globules) which occurred singly, in clusters, or in chains, which had rigid cell walls, and were either motile or nonmotile, have been redistributed between the two orders Pseudomonadales and Eubacteriales. The former includes all the bacteria with polar flagella and the latter those with peritrichous flagella. Since all the nonmotile species lack flagella their classification has been based on a comparison of their physiological characteristics with those of the respective orders. The species of the Pseudomonadales have a more marked dependence on oxygen than the species of the Eubacteriales.

The division is not simple. In some of the Eubacteriales, for example the genus *Chromobacterium*, the distribution of the flagella is affected by the environment, and both polar and peritrichous forms may be produced. Similarly, some polar flagellates have a distinctly fermentative metabolism which does not require oxygen. In such cases the balance of evidence is used to decide into which category the organism is to be placed. *See* BEGGIATOALES; PSEUDOMONADALES.



1. *Azotobacter* (Azotobacteraceae)
2. *Rhizobium* (Rhizobiaceae)
3. *Alcaligenes* (Achromobacteraceae)
4. *Escherichia* (Enterobacteriaceae)
5. *Klebsiella* (Enterobacteriaceae)
6. *Pasteurella* (Brucellaceae)
7. *Fusobacterium* (Bacteroidaceae)
8. *Sphaerophorus* (Bacteroidaceae)
9. *Micrococcus* (Micrococcaceae)
10. *Gaffkya* (Micrococcaceae)
11. *Sarcina* (Micrococcaceae)
12. *Neisseria* (Neisseriaceae)
13. *Veillonella* (Neisseriaceae)
14. *Kurthia* (Brevibacteriaceae)
15. *Diplococcus* (Lactobacillaceae)
16. *Streptococcus* (Lactobacillaceae)
17. *Leuconostoc* (Lactobacillaceae)
18. *Lactobacillus* (Lactobacillaceae)
19. *Butyribacterium* (Propionibacteriaceae)
20. *Corynebacterium* (Corynebacteriaceae)
21. *Arthrobacter* (Corynebacteriaceae)
22. *Bacillus* (Bacillaceae)
23. *Clostridium* (Bacillaceae)

Representative genera of the Eubacteriales; the family is given in parentheses.

The original conception of the Eubacteriales as "true" bacteria has thus been lost and the order has adopted a new meaning as indicated previously.

Families. On the basis of the shape (morphology) and internal structure of the cells, their reaction to stains (cytology), their sensitivity to oxygen, and their food requirements (physiology), the order has been subdivided into the following 13 families: Azotobacteraceae, Rhizobiaceae, Achromobacteraceae, Enterobacteriaceae, Brucellaceae, Bacteroidaceae, Micrococcaceae, Neisseriaceae, Brevibacteriaceae, Lactobacillaceae, Propionibacteriaceae, Corynebacteriaceae, and Bacillaceae. Representative genera are shown in the illustration.

Division on morphology is perhaps the most important since the shape and internal structure of the cell, apart from being the first characteristics to meet the eye, probably reflect the genetic constitution more than any other character. In the Eubacteriales there are only two distinct morphological forms: the rod-shaped cell and the spherical cell or coccus. The families Micrococcaceae and Neisseriaceae comprise essentially spherical cells, the former staining gram-positive and the latter gram-negative. There are, however, other genera of gram-positive cocci which have been placed in the tribe Streptococceae in the family Lactobacillaceae, since some taxonomists consider the physiological property of lactic acid production more significant than the shape of the cells. Others may well question this arrangement. *See* GRAM'S STAIN.

The rod-shaped organisms are divided primarily on their reaction to Gram's stain. The Azotobacteraceae, Achromobacteraceae, Rhizobiaceae, Enterobacteriaceae, Brucellaceae, and Bacteroidaceae are gram-negative and the Brevibacteriaceae, Lactobacillaceae, Propionibacteriaceae, Corynebacteriaceae, and Bacillaceae, gram-positive.

Of the gram-negative families, the Bacteroidaceae do not grow in the presence of oxygen (anaerobic), except for the genus *Streptobacillus*, which may well belong elsewhere. Of the aerobic organisms the Azotobacteraceae fix atmospheric nitrogen nonsymbiotically; the Rhizobiaceae form nodules or tuberculous lesions on leguminous and other plants. Within this family, the genus *Rhizobium* forms a symbiotic relationship with the legume and fixes atmospheric nitrogen, while the genus *Agrobacterium* is parasitic. Inclusion of the genus *Chromobacterium* is apparently based on variation in the form of flagellation.

The Enterobacteriaceae include those organisms found in the alimentary canal of man or animals under normal and pathological conditions, and forms closely related to these organisms in their general physiological and antigenic characters.

The Brucellaceae are a group of animal parasites which have rather fastidious nutritional requirements resulting from a long and intimate association with animal hosts to the extent that they have become dependent to a large extent on preformed animal foods.

The Achromobacteraceae are a group of organisms found mainly in soil and water.

The separation of families of the gram-positive rods is based on morphology and physiological properties. The Bacillaceae produce endospores. The Lactobacillaceae produce acid, principally lactic, in considerable quantities from sugars. The Propionibacteriaceae produce propionic and butyric acids or ethyl alcohol and frequently some carbon dioxide. The remaining families, the Brevibacteriaceae and Corynebacteriaceae, are difficult to distinguish. In the former the shape of the cells is usually uniform and constant whereas in some genera of the latter bizarre pleomorphic forms are frequently produced. Taken collectively, neither family has any pronounced fermentative activity. For further discussion, see separate articles on the individual families. *See also* BACTERIA, TAXONOMY OF. [V.B.D.S.]

Eucalyptus

A large genus of evergreen trees, with about 600 species, belonging to the myrtle family (Myrtaceae), and occurring in Australia, New Guinea, and the Sunda Islands. It has been widely planted as an ornamental tree. Some species grow to be very large trees. *Eucalyptus regnans* attains a height of 326 ft and a diameter of 25 ft. Some species are valuable for fuel and lumber. The leaves of *Eucalyptus globulus* (blue gum) contain an essential oil used in medicines and perfumes. *See* MYRTALES. [P.D.S.]



Eucalyptus globulus showing spray of mature foliage and two leaves of sucker foliage. (L. H. Bailey, The Standard Cyclopedia of Horticulture, vol. 1, Macmillan, 1937)

Eucarida

A large and economically important superorder of higher Crustacea (Malacostraca), including shrimps, lobsters, hermit crabs, and crabs. They have the following characteristics: the shell is

fused dorsally with all the thoracic segments, the eyes are on movable stalks, and there is no lacinia mobilis on the mandible in the adult. Eggs are usually attached to the abdominal appendages which may have an appendix interna. The liver is much ramified; the heart is short and thoracic. Spermatozoa are spherical, often with radiating processes, and development occurs as a rule with metamorphosis and a free swimming nauplius stage in primitive forms.

It comprises two orders, the Euphausiacea and Decapoda. The Euphausiacea, a small oceanic group important as food of whales, retains the caridoid facies. The exopodite of the maxilla is minute and the exposed gills are in a single series. None of the thoracic limbs are termed maxillipeds, but all may be specialized to form a filter feeding mechanism. Their similarity is thus secondary rather than primitive. The young hatch as nauplii. No fossil forms are known. In Decapoda the caridoid facies may be retained as in the shrimps but is often greatly modified, culminating in the crablike carcinoid facies. The exopodite of the maxilla is large. The gills are protected by the shell and are generally in several series. They are rarely absent; however, *Lucifer* is an example in which they are lacking. Three pairs of thoracic limbs are specialized as maxillipeds, the others are pereiopods. Generally the young hatch at a later stage than the nauplius. Direct development occurs in freshwater crayfish, river crabs, and a few marine crabs. Numerous fossil forms from the Triassic and Jurassic onwards are known *See* DECAPODA (CRUSTACEA); EUPHAUSIACEA; MALACOSTRACA [I.GO.]

Eucestoda

A subclass of the Cestoda or tapeworms. This term is used frequently to distinguish the true tapeworms from the tapewormlike forms which comprise the subclass Cestodaria. In those classification schemes in which the Cestodaria are given class status, the true tapeworms comprise the class Cestoda. *See* CESTODA; CESTODARIA. [C.B.C.]

Eucoelomata

That large sector of the animal kingdom in which there is a true coelom or body cavity, ordinarily lined by a layer of cells, the peritoneum. The coelom is the space inside the body containing the digestive tract and other internal organs or viscera. All phyla above the Aschelminthes, from Bryozoa to Chordata, are included in Eucoelomata. The coelom is large in Echinodermata, Annelida, and Chordata but much reduced in Mollusca and Arthropoda. *See* ACOELOMATA; PSEUDOCOELOMATA. [T.I.S.]

Euechinoidea

A subclass of Echinoidea, distinguished by relative stability of the ambulacra and interambulacra, each comprising two columns of plates. J. Durham and R. Melville (1957) include 14 orders in the Echinoidea. These are the Diadematoida, Echinothurioida, Pygasteroida, Hemicidaroida, Phymosomatoida, Arbacioida, Temnopleuroida, Echinoida, Holectypoida, Clypeasteroida, Nucleolitoida, Cassiduloida, Holasteroida, and Spatangoida. *See* DIADEMATACEA; ECHINACEA. [H.B.F.]

Eugenics

The study of factors that influence the hereditary qualities of future generations. It may be thought of as both a science and a social movement. Eugenics proposes to improve man's future by increasing the number of children produced by persons who are, by some definition, superior and by reducing the number produced by persons who are physically or mentally deficient. Attempts to encourage larger families from superior parents are called positive eugenics, attempts to reduce the number of children from defective parents negative eugenics.

The eugenics movement was started in the 1880s by Sir Francis Galton. Galton was much influenced by the evolution theory of his cousin, Charles Darwin, and was one of the first to argue that man's future evolution is in his own hands In proposing the word eugenics he suggested that man is now able to replace the cruel and blundering process of natural selection with processes under his own control that are no less effective and much more merciful.

Man's genetic future. In considering man's genetic future the eugenist sees two causes for concern: a differential birth rate, and the decreasing effectiveness of natural selection. A differential birth rate is shown in many studies made in the United States and Great Britain in the period from 1920 to 1940. These data show a distinct tendency for parents of higher education, of more intellectually demanding occupations, and of greater attainments to have a smaller number of children than those at the other end of the scale To whatever extent these attainments are genetically determined, the genetic potential of future generations is being weakened by these trends. However, it is likely that such differential birth rates are less pronounced now than formerly, the strongly differential rates in the earlier years being the result of unequal availability of birth control information at that time

The decreasing effectiveness of natural selection is a consequence of the increase in living standards and the advances in medicine. Diseases that once were fatal are now curable. Some of these diseases are genetic in origin, which means that deleterious genes which once were eliminated by the premature death of the person carrying them now are transmitted to future generations. Thus, one of the consequences of an improved environment is an increase in the number of harmful genes accumulating in the population. Since new harmful mutations are occurring in every generation because of mutation, these are added to those already present.

Geneticists differ in their views as to the seriousness and urgency of these problems. In the

eugenists have given consideration mostly to the first question, but at the present time, particularly with the emphasis on the possible increase in mutation as a consequence of nuclear radiations, the second factor is being more seriously considered.

Negative and positive genetics. There have been a number of proposals regarding negative eugenics in the United States. Over half of the states have laws providing for sterilization of persons with severe mental or physical defects. Segregation of persons in public institutions also has the effect of reducing the number of births. Finally there is the advocacy of greater use of contraceptives by persons likely to transmit hereditary weaknesses. Positive eugenics has consisted largely of exhortations for the well endowed to have more children.

Such eugenic proposals have not had complete acceptance. There are objections on religious grounds. There are those who think that eugenic laws, particularly if they involve compulsory sterilization, are an unwarranted infringement on basic human rights. A different reservation comes from those who point out that knowledge of human heredity is not very complete and that prediction of the attributes of children from their parents is not very exact. These persons point out that the rate of genetic deterioration at present is surely extremely slow relative to the rate of increase in the knowledge of genetics, and that concrete eugenic measures might well await increased knowledge that would reduce the number of mistakes.

Contemporary discussions of eugenics place much more emphasis on voluntary control over reproduction than on any compulsory population policies. Recent advances in genetics have shown new ways of detecting persons who carry hidden genes for various mental and physical conditions. Some medical schools now have heredity clinics which offer genetic counseling to prospective parents.

Euthenics. Future improvement of the human race depends not only on improving its heredity, but also on improvements in environment. The improvement of the future of man by changing his environment is known as euthenics. [J.F.C.]

Bibliography: F. Osborn, *Preface to Eugenics*, rev. ed., 1951.

Euglena

A common form of one-celled life, variously classified as a member of the class Mastigophora, also called Flagellata, phylum Protozoa, and sometimes as green algae (*see* ALGAE). *Euglena* thrives in stagnant fresh water where it often forms a green scum over the surface, or conspicuous green spots [illegible]

[illegible]

[illegible]thesis by means of which it produces part of its food. Primarily, however, food appears to be absorbed through the body wall from the water in which it lives.

Typically reproduction is by means of longitudinal fission, although a primitive form of sexual reproduction is known in at least one species. Under certain conditions, for example, hot weather or strong light, cysts are produced, each containing 16–32 small daughter euglenoids formed by longitudinal fission. Like the cysts of other simple forms, these are highly resistant stages that will resume activity under favorable environmental conditions. *See* MASTIGOPHORA. [J.D.B.]

Euglenida

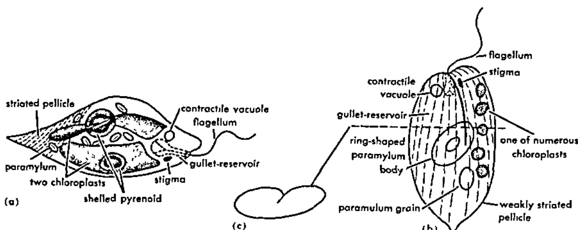
An order of the class Phytomastigophorea. This order of protozoans, also known as Euglenoidina, includes the largest green noncolonial flagellates, *Euglena ehrenbergii*, which are 400 μ long. Many of the colorless members are also large. They have one or two equal or subequal flagella. There are relatively few genera of Euglenida, even if doubtful ones such as *Clorachne* and *Ottonia* are included. The fresh-water genera *Euglena*, *Lepocinclis*, *Phacus* (see illustration), and *Trachelomonas* have many species, as does the colorless genus *Petalomonas*. Others have few species, including the marine genus *Eutreptia*. *Calkinsia* is yellow, some euglenas are red at times, and all others are green or colorless, although *Trachelomonas* tests are frequently colored by iron.

Morphology. Euglenids generally show an anteroposterior elongation; however, *Trachelomonas volvocina* is practically spherical, as shown by young individuals, and *Euglena acus* is extremely needlelike. Most free-swimming members are round in cross section, but species of *Phacus* are flattened, as are *Petalomonas*, *Anisonema*, *Triangulomonas*, and others. Many are enclosed in a thick pellicle which is sculptured into keels (*Phacus* spp., *Pleotia*, and *Tropidoscyphus*); others such as *Euglena oxyuris* and *Phacus pyrum* shows strong torsion of the pellicle. This pellicle is sometimes ornamented with striae or small wartlike protuberances. It prevents distortion of the body according to its degree of thickening, but species which have a very thin pellicle (*Astasia*, *Peranema*, *Heteronema* and some euglenas) shorten on their axis, then elongate. Such contortions are termed euglenoid or metabolic movements. There are sometimes fine openings in the pellicle through which mucus threads are extruded under certain conditions.

Colony formation is restricted to *Colacium*, in which a flagellated euglenoid cell usually attaches to a copepod by its anterior end and develops into an arboroid structure of several nonflagellated cells. Tests or shells are common; *Klebsiella* is a free-swimming marine form in an urn, *Ascoglena* a fresh-water epiphyte in an urn. In *Trachelomonas* and *Strombomonas* the tests vary widely in shape and ornamentation [illegible] through

[illegible] ve two [illegible] starch- [illegible] stant shape [illegible] en species. [illegible] and color- [illegible] large and lamellated.

The second characteristic is the gullet-reservoir system. An anterior mouth opens into a short



(a) *Euglena pisciformis.* (b) *Phacus stokesii.* (c) Cross section of cell.

gullet which widens into a reservoir. The flagella emerge from its floor, and at one side a contractile vacuole pulses, except in marine species. This system is present in some Chloromonadida (*Gonyostomum semen*) and is approximated by certain Cryptomonadida, but all euglenoids possess it. The mouth may serve to ingest food (*Peranema trichophorum*), but in some cases does not. Green forms possess a stigma near the reservoir, and a thickening of one root of the bifurcated flagellum has been found in some uniflagellate species.

Flagella and chromatophores. In most green species the flagellum is long, uniform, and movable throughout. *Euglena mutabilis* seems to lack a flagellum, while *E. spirogyra* has a short flagellum; *Calkinsia* has a long, uniform, anterior one, movable throughout, and a shorter, fine, trailing one. *Peranema* has a long tapering flagellum, vibratile at its tip during steady progression, and a very short, fine, posterior one. *Anisonema acinus* has a rather short, fine, anterior flagellum and a tapering, long, and very heavy posterior one. Chromatophores in the green members may be circular, stellate, or irregular disks, or variously shaped plates and short bands. The chromatophores are reticulate plastids containing bright green pigments closely related to, if not identical with, the chlorophylls of higher plants. The chromatophores of Euglenida never contain other masking pigments. Often pyrenoids are associated with them. These are probably plastids concerned with the elaboration of paramylum, although paramylum is formed in species lacking pyrenoids. In *Euglena pisciformis* both chromatophores have a central pyrenoid which has a shell of paramylum. Pyrenoids stain deeply with hematoxylin and may be proteinaceous. Besides paramylum, fats occur, and glycogen has been reported.

Nuclei. Nuclei are large, vesicular, and contain one or more endosomes which are deeply staining chromatoid bodies. These are surrounded by dispersed chromatin granules. In division, a specific number of chromosomes is formed, but a spindle is lacking, and chromosomes are oriented around and parallel to the drawn-out endosomal mass. Cell division usually occurs in the active state, although *Euglena gracilis* divides while encysted. Palmella stages have been reported but are not common. Sexual processes are unknown. Daily isolations of several clones of *Entosiphon sulcatum* for many months indicated no depression of the division rate, and the usual type of mitotic division was sufficient to maintain species vitality.

Nutrition. Nutrition varies greatly in Euglenida. Green members are holophytic or autotrophic, but there is evidence that some of them absorb soluble organic matter with no evidence of ingestion of solid matter. It seems questionable that more than a few green members are strict autotrophs which do not require organic carbon and nitrogen but form carbohydrates and proteins from carbon dioxide and inorganic salts. Certainly the colorless forms are heterotrophs (dependent on organic food), and some of the green ones readily assume a heterotrophic existence, if they ever were autotrophic. Thus *Euglena gracilis* and species of *Phacus* become colorless in an organic medium in the dark, whereas *Trachelomonas reticulata* is devoid of chlorophyll. *Peranema* ingests large food bodies; however, most of the colorless forms are saprozoic rather than holozoic and are readily cultivated in soluble organic media. They also vary widely in oxygen requirements; the same species will occur in well-oxygenated or anaerobic situations. In a rich organic medium and under a film of oil, *Entosiphon* will thrive, but often forms somatellalike masses, in which, however, cell division does not become complete. Parasitism is unknown in the class, although *Euglenomorpha* occurs in the intestine of tadpoles, where it is probably commensal.

Ecology. In habitat Euglenida are widespread. Green members occur mostly in fresh water and frequently in such numbers as to form blooms. This is especially true for pools contaminated by cattle droppings. *Strombomonas* in the Ohio Valley of the United States is a potamophile form, *Trachelomonas* a lake or pond form. *Eutreptia* blooms densely in marine bays and estuaries, and the interface in marine waters is a rich sour

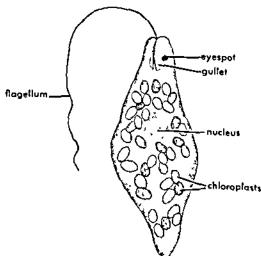
the colorless creeping members. Iron seepages and coal-mine seepages containing sulfuric acid are often bright green with *Euglena mutabilis* even at pH values of 1.5, whereas cedar swamps have blooms of *Euglena polymorpha*. Unlike Phytomonadida, Euglenida appear in warmer waters but do not seem to occur in hot springs. They abound in citrus-waste lagoons and in later stages of sewage-treatment waters, an indication of their partially saprozoic nutrition.

Relationships within the class are not well defined at present; each classification within recent years has departed from previous ones. Organisms are readily assigned to the Euglenida on a structural basis, because the class is quite specialized. However, intraclass relationships are difficult to assign. *See* MASTIGOPHORA. [J.B.L.]

Euglenophyta

A group of microscopic, one-celled, chiefly aquatic organisms. Because they possess the characteristic green plant pigment, chlorophyll, they are classified as a phylum of the subkingdom Thallophyta. Zoologists, however, classify them as animals in the phylum Protozoa, chiefly because they have a gullet through which solid food particles may be taken into the body. A typical euglenoid has a spindle-shaped body, usually naked and bearing one or two (rarely three) projecting flagella, or protoplasmic tails, the rhythmic movements of which rapidly propel the organism through water. With the aid of eir chlorophyll, borne in definite protoplasmic dies called chloroplasts, euglenoids carry on photosynthesis, the manufacture of carbohydrate food from carbon dioxide and water using the energy of sunlight. In this respect, they resemble all other green plants (*see* PHOTOSYNTHESIS).

Reproduction is chiefly by lengthwise, asexual division of one body into two. Euglenoids are found



Euglena with anterior flagellum, gullet, and eyespot; ntral nucleus; c numerous chloroplasts (greatly nified). (H. J. and O. Tippo, *College Botany*, Holt, 1954)

in great numbers in both fresh and ocean waters and in moist soil. Many biologists regard them as closely related to the original stock of primitive organisms from which evolved both the plant kingdom and the animal kingdom. The chief economic value of euglenoids lies in their importance as a major source of food for fish and other aquatic animals. *See* ALGAE; PROTOZOA. [H.J.F.]

Bibliography: *See* THALLOPHYTA.

Eugregarinida

An order of the subclass Gregarinidia. These protozoans are common parasites of invertebrates such as the arthropods and annelids. Only sexual reproduction occurs in the life history of these animals; schizogony is lacking. This order is divided into the suborders Cephalina and Acephalina, based on the body of the sporont or trophozoite which in the cephaline gregarines is divided by a septum into two parts, the protomerite and deutomerite. The body of the acephalines is not divided. In North America, grasshoppers of the genus *Melanoplus* are commonly the host for the cephaline *Gregarina rigida* which inhabits the lumen of the gut. [E.R.BE.]

Eulamellibranchia

The largest subclass in the class Pelecypoda, the bivalve mollusks. Eulamellibranchia contains many food species, such as *Mercenaria mercenaria* L., the quahog clam found along the East Coast of North America; *Cerastoderma edule* L., the cockle of western Europe; *Ensis directus* Say., the razor clam of eastern North America; and many other edible clams throughout the world.

Morphology. The shell hinge is heterodont with few and varied teeth and sockets. The gills are leaflike, with permanent junctions between the gill filaments. The anterior and posterior adductor muscles are about equal in size. The mantle is partially to completely closed along the ventral margin and the siphons are well developed. Eulamellibranchia are world-wide in distribution and occur in marine and fresh-water environments.

Ecology. The family Unionidae, the fresh-water mussels, is nearly world-wide in distribution. The larvae, known as glochidia, are parasitic on fish, attaching themselves to the gills and fins where they remain encysted for various periods of time, depending upon the species. Various species in the genera *Unio* and *Anodonta* are widely distributed in Europe and the Near East, and many genera such as *Elliptio*, *Anodonta*, *Crenodonta*, *Lampsilis*, and *Quadrula* occur in North America, particularly in the Mississippi-Ohio river system.

Many species of fresh-water mussels which have a white and lustrous nacre are important as a source of pearl buttons, and occasionally they produce valuable commercial pearls.

This subclass also contains the families Teredinidae and Pholadidae. Various species in these families cause extensive damage to all kinds of wooden installations, such as ships and wharves, by boring into the wood and even into cables and poor grades of cement.

The damage done by *Teredo navalis* L., the northern shipworm, exceeds many times in value the money derived from all of the edible species in this order.

This subclass also contains the family Tridacnidae, the giant clams of the tropical Pacific. Large examples of *Tridacna gigas* L. exceed 3 ft in length and weigh as much as 600 lb. They live embedded in coral reefs from the intertidal zone to depths of a few fathoms. They add to their food supply by growing a huge amount of unicellular algae in their tissues which they utilize when other food is scarce. The family Corbiculidae is rich in species, particularly in the Orient. They live in fresh water and are a very important source of food. *See* PELECYPODA; SHIPWORM. [W.J.C.]

Euler angles

Three angular parameters that specify the orientation of a body with respect to reference axes. They are used for describing rotating systems like gyroscopes, tops, molecules, and nonspherical nuclei. They are not symmetrical in the three angles but are simpler to use than other rotational parameters.

Unfortunately, different definitions of Euler's angles are used, and therefore it is confusing to compare equations in different references The definition given here is the majority convention according to H. Margenau and G. Murphy

Let $OXYZ$ be a right-handed cartesian (right-angled) set of fixed coordinate axes and $Oxyz$ a set attached to the rotating body

The orientation of $Oxyz$ can be produced by three successive rotations about the fixed axes starting with $Oxyz$ parallel to $OXYZ$ Rotate through (1) the angle ψ counterclockwise about OZ, (2) the angle θ counterclockwise about OX, and (3) the angle ϕ counterclockwise about OZ again. The line of intersection OK of the xy and XY planes is called the line of nodes.

Denote a rotation about OZ, for example, by Z (angle). Then the complete rotation is, symbolically,

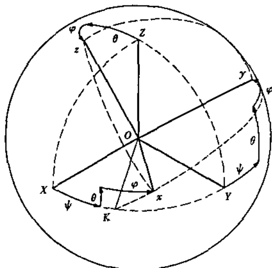
$$R(\psi,\theta,\phi) = Z(\phi)X(\theta)Z(\psi)$$

where the rightmost operation is done first.

A point P will have coordinates (x,y,z) with respect to the body axes and (X,Y,Z) with respect to the fixed axes. These are related by the linear equations:

$$\begin{aligned} x &= X\cos(x,X) + Y\cos(x,Y) + Z\cos(x,Z) \\ y &= X\cos(y,X) + Y\cos(y,Y) + Z\cos(y,Z) \\ z &= X\cos(z,X) + Y\cos(z,Y) + Z\cos(z,Z) \end{aligned}$$

where (x,X) is the angle between the axes Ox and OX, etc. The nine direction cosines are expressed



Euler angles. The successive movements of the axes on a unit sphere described in the text are shown by arrows. The complete rotation may also be obtained by a different sequence of rotations, namely, first through φ about OZ, then through θ about the displaced x axis (which is OK), then through ψ about OZ.

in terms of the three Euler angles by the table at the bottom of the page.

Inspection of the figure makes it apparent that no operation in $R(\psi, \theta, \phi)$ can be replaced by a combination of the other two. Therefore, three parameters are needed to specify the orientation, and the amounts of the angles are unique (barring additional 360° rotations) In dynamical problems of rotating bodies, ψ, θ, and ϕ can be used as independent angular coordinates.

Molecules and nuclei undergo oscillatory changes in shape while rotating. The body axes apply to the average shape. For another set of rotational parameters *see* CAYLEY-KLEIN PARAMETERS. [B.C.]

Bibliography: H Goldstein, *Classical Mechanics*, 1950; H. Margenau and G M. Murphy, *The Mathematics of Physics and Chemistry*, 2d ed., 1956.

Euler's equations of motion

A set of three differential equations expressing relations between the force moments, angular velocities, and angular accelerations of a rotating rigid body. They are equations of motion in the usual dynamical sense, having the form

$$\begin{aligned} I_1(d\omega_1/dt) + (I_3 - I_2)\omega_2\omega_3 &= M_1 \\ I_2(d\omega_2/dt) + (I_1 - I_3)\omega_3\omega_1 &= M_2 \\ I_3(d\omega_3/dt) + (I_2 - I_1)\omega_1\omega_2 &= M_3 \end{aligned}$$

	X	Y	Z
x	$\cos\psi\cos\phi - \sin\psi\sin\phi\cos\theta$	$\cos\psi\sin\phi + \sin\psi\cos\phi\cos\theta$	$\sin\psi\sin\theta$
y	$-\sin\psi\cos\phi - \cos\psi\sin\phi\cos\theta$	$-\sin\psi\sin\phi + \cos\psi\cos\phi\cos\theta$	$\cos\psi\sin\theta$
z	$\sin\phi\sin\theta$	$-\cos\phi\sin\theta$	$\cos\theta$

The formulation employs as coordinate axes the three principal axes of rotational inertia of the body which can rotate about a body-fixed point, which is the center of mass if constraints are absent. These reference axes, which form a right-hand set, are indicated by subscripts 1, 2, and 3 in the equations, where I_1, I_2, and I_3 represent the principal moments of inertia; ω_1, ω_2, and ω_3 the angular velocities about the axes; M_1, M_2, and M_3 the corresponding force moments; and t the time.

In the general case these equations cannot be integrated, but solution is possible in special cases. Soluble problems of interest include that in which force moments are absent, the resulting complex behavior being called Poinsot motion, and that in which two of the principal moments of inertia are identical and only one force moment is present. The latter case includes spinning tops and gyroscopes. *See* POINSOT'S METHOD; RIGID BODY DYNAMICS. [R.A.F.]

Bibliography: H. C. Corben and P. Stehle, *Classical Mechanics*, 1950; J. C. Slater and N. H. Frank, *Mechanics*, 1947; E. T. Whittaker, *A Treatise on the Dynamics of Particles and Rigid Bodies*, 1904.

Euler's momentum theorem

A principle of fluid mechanics which states that momentum of the particles in a moving frictionless, or inviscid, fluid is conserved. This theorem is expressed by Euler's hydrodynamical equations, a set of nonlinear partial differential equations. When viscous shear forces are included, the momentum equations are called the Navier-Stokes equations. The Eulerian method of viewing fluid motion is to consider it as a velocity-pressure field; that is, the velocity of the fluid particles is considered as a vector function of position in space and time, $\mathbf{v} = \mathbf{v}(\mathbf{r},t)$, and the pressure as a scalar function of position and time, $p = p(\mathbf{r},t)$ where $\mathbf{r}$ is the position vector locating a point in space and t is time. Euler's equations are essentially a restatement of Newton's single-particle law, force equals mass times acceleration, adapted to the many-particle continuum concept of fluid motion. In such a flowing continuum the acceleration of a particle of fluid at a given point in the fluid is given by

$$\mathbf{a} = d\mathbf{v}/dt = \partial\mathbf{v}/\partial t + (\mathbf{v}\cdot\nabla)\mathbf{v}$$

where $(\partial\mathbf{v}/\partial t)\,dt$ represents the velocity change in the time interval dt at a fixed point and $(\mathbf{v}\,dt\cdot\nabla)\mathbf{v}$ represents the velocity change of the particle in traveling the distance $\mathbf{v}\,dt$ from one point in the fluid to another point during time dt. Consider a volume V in the fluid with a surface S and let $\mathbf{n}$ be a unit vector normal to the surface, positive outward. One of the forces acting on this volume is the resultant of the pressure forces

$$-\iint_S p\mathbf{n}\,dS$$

which, by the analog to Gauss' divergence theorem, is equivalent to

$$-\iiint_V \nabla p\,dV$$

Thus, the negative of pressure gradient $-\nabla p$ may be regarded as a force per unit volume acting on a fluid particle. There may also be body forces such as gravity and electromagnetic forces (if the fluid is electrically conducting). Let the amount of body force per unit mass be $\mathbf{F}$, then the force per unit volume is $\rho\mathbf{F}$ where ρ is the density. Euler's equations (as one vector equation) then become

$$\rho[(\partial\mathbf{v}/\partial t) + (\mathbf{v}\cdot\nabla)\mathbf{v}] = -\nabla p + \rho\mathbf{F}$$

where the left-hand side is (mass per unit volume) times acceleration and the right-hand side represents the forces per unit volume. In rectangular Cartesian coordinates (x,y,z) with velocity components $(u,v,w,)$ and body force components (X,Y,Z), the equations are

$$\rho\left[\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z}\right] = -\frac{\partial p}{\partial x} + \rho X$$

$$\rho\left[\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z}\right] = -\frac{\partial p}{\partial y} + \rho Y$$

$$\rho\left[\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z}\right] = -\frac{\partial p}{\partial z} + \rho Z$$

An integral of these equations can be found for flows starting from rest or uniform motion (*see* KELVIN'S CIRCULATION THEOREM). In this way the problem of determining the velocity field is reduced to purely kinematic considerations. *See* BERNOULLI'S THEOREM; GAS DYNAMICS; LAPLACE'S IRROTATIONAL MOTION; NAVIER-STOKES EQUATION. [A.E.BR.]

Eumalacostraca

A series of the class Crustacea which contains the superorders Syncarida, Peracarida, Eucarida, Hoplocarida, and Pancarida. These higher crustaceans are shrimplike in appearance and have several invariable characteristics. The thorax has 8 segments while the abdomen has 6 segments in addition to the telson or tail fan. The last pair of pleopods has a specialized form and function. The uropods of the Syncarida and Mysidacea have statocysts. In many orders of Eumalacostraca the first or second pleopods in the male are modified for sexual purposes. The stomach is divided into two chambers, the anterior cardiac portion, containing the gastric mill which mechanically grinds the food, and the posterior pyloric portion, in which chemical digestion is initiated. Characteristic larval stages are the nauplius, contained within the egg, and the mysis and zoaea, which are free-swimming. These organisms are free-living or parasitic. The parasites show a great adaptability to their hosts which are fish or other crustaceans. The majority of the species are marine and many, except the Hoplocarida, show

adaptive tendencies to brackish and fresh-water environments. The oviducts open on the sixth thoracomere while the sperm ducts open on the eighth.

Systematics and phylogeny. Five superorders comprise the Eumalacostraca. The Syncarida have complete segmentation, lack a carapace, and are the most primitive of the eumalacostracans. They *are restricted to special biotopes.*

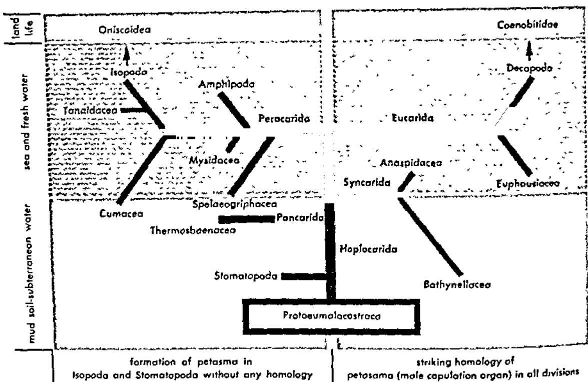
Peracarida have the first thoracic segment united with the head. The carapace may be large, small, or absent (secondarily reduced). Some thoracic segments are free from the carapace. The cephalothorax is usually larger than the abdomen. The eyes may be pedunculate, sessile, or absent.

Eucarida *show a well-developed cephalothorax.* They have complicated life cycles which include larval metamorphosis passing through all characteristic stages. The eyes are stalked, and the heart and gills are located in the thoracic region. Eggs are carried on the inner branch of the pleopods.

The Hoplocarida comprise a group of about 200 *species of eumalacostracans with one order, the* Stomatopoda, and one family, the Squillidae. The body shows both primitive characteristics in the nervous system and elongated heart and specialized characters in the maxillipeds, first pleopods, and antennae. The abdomen is very large and terminates in a large telson and wide uropods. The cara-

Characteristics of the orders of the Eumalacostraca

Classification	Striking characters	Habitat	Pairs of maxillipeds	Development	Eggs	Carapace
Superorder Syncarida						
Order Anaspidacea	*Primitive biramous thoracic limbs*	*Fresh waters of Tasmania*	*None*	*Direct*	*Laid free*	*Lacking*
Order Bathynellacea	Furca	Subterranean waters; cosmopolitan	None	Direct	Laid free	Lacking
Superorder Peracarida						
Order Isopoda	Body dorsoventrally compressed	Marine; fresh water, terrestrial	1	Direct	Incubated in a brood pouch composed of a varying number of lamellae or oostegites	Lacking
Order Tanaidacea	Claw-bearing first pair of legs	Marine; sometimes fresh water	1	Direct		Reduced
Order Amphipoda	Body laterally compressed	Marine; fresh water	1	Direct		Lacking
Order Mysidacea	Antennae with squama	Marine	1, possibly 2	Direct		Present
Order Cumacea	Eyes coalesced	Marine; mud	2, possibly 3	Direct		Present
Order Spelaeogriphacea	Blind	Cavernicolous	1	Direct		Present
Superorder Pancarida						
Order Thermosbaenacea	Blind	Warm springs	1	Direct	Incubated in a brood pouch composed of a varying number of lamellae or oostegites	Present
Superorder Eucarida						
Order Euphausiacea	Biramous peraeoand pleopods	Marine plankton	None	With metamorphosis	Generally laid free; sometimes cemented to thorax of female	Present
Order Decapoda	10 pairs walking legs	Marine; fresh water; terrestrial	3	Generally with metamorphosis; sometimes abbreviated or direct	Sometimes freely shed; generally attached to pleopods of female	Present
Superorder Hoplocarida						
Order Stomatopoda	Raptorial arms (especially the second pair of maxillipeds)	Marine; mud	5	Metamorphosis with distinct stages	Laid free	Present



Phylogenetic tree of Eumalacostraca.

pace is small Five pairs of raptatorial appendages are present of which the second is especially modified. These animals are exclusively marine.

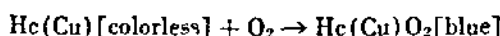
Pancarida are eumalacostracans of small size The single order, Thermosbaenacea, contains four species The body is cylindrical, and both eggs and embryos are carried in a dorsal marsupium formed by the carapace.

Anatomy and physiology. While the external and internal organs of the most important families are well known, certain aspects of physiology, such as endocrinology, function of the nervous system, metabolism in general, and the chemical processes of metamorphosis, are still largely unknown.

The labrum contains salivary glands, whose secretions are rich in mucus. The esophagus leads into a stomach divided into a cardial chamber, containing the gastric mill, and a pyloric chamber, where products of the digestive glands, especially enzymes such as amylases, proteases, and lipases, are discharged. The short midgut follows the stomach and terminates in the rectum. From the developmental point of view, there are three parts to the digestive tract—the stomodeum (ectodermal), mesodeum (entodermal), and proctodeum (ectodermal). Both the stomodeum and proctodeum are lined with chitin. There are four layers comprising the digestive tube. The inner membrane, or intima (chitinous), is derived from the cuticula of the stomodeum and proctodeum. The epithelium in the mesodeum is connected to the hypodermis of the skin, but is exposed in the midgut as the true digestive membrane. It functions in the resorption of amino acids, fatty acids, monosaccharides, and glycerine. The muscularis consists of longitudinal and circular layers, and the serosa is connective tissue on the outer surface of the digestive tract. Sometimes there is a very small midgut with diverticuli or branched ceca which are of exceptional extension and differentiation. These are involved in the digestive process. The hepatopancreas is located in the forepart of the mesodeum and has digestive and excretory functions. The different parts of the intestine are separated from each other by valves which control food passage. The proctodeum functions in the elimination of excrements.

Circulatory and respiratory system. Blood courses throughout the body cavity which is divided into several chambers, separated from one another by special partitions. The heart, surrounded by the pericardial sinus, is a contractile dorsal vessel which maintains the flow of blood into the arteries by its pumping movements. The heart sizes vary greatly within the Eumalacostraca. They are reduced in the Syncarida, saclike in many Decapoda (crabs), and occupy the entire dorsal region of the thorax and abdomen in Hoplocarida. They are provided with pairs of valves, which regulate the flow of blood entering through ostia from the pericardium. There is a direct correlation between heart size and gill position in all eumalacostracan groups. Cutaneous respiration occurs but is insufficient. The exopodites of the thoracic limbs and pleopods of the Isopoda are modified as gills. The Decapoda possess respiratory chambers with pleuriramous gills. There are several adaptations to

land life as in *Birgus latro*, the coconut crab, which has lunglike spaces formed by the branchiostegites which cover the gills. All higher Eumalacostraca are provided with supplementary arteries of various sizes and extension, such as the antennal, subneural, and hepatic arteries. In general, the blood is colorless, but several species having high oxygen requirements have respiratory pigments dissolved in the blood plasma. The most important pigment is hemocyanin, a copper-porphyrin, which unites with oxygen in the following manner and with an accompanying change in color.



$$Hc(Cu)[\text{colorless}] + O_2 \rightarrow Hc(Cu)O_2[\text{blue}]$$

The amoebocytes, white cells with pseudopods, are characteristic. They are phagocytic and necessary for blood coagulation.

Two essential glands occur, the maxillary and antennary. During the larval stages there is a functional maxillary gland which is later superseded by the antennal gland, the so-called green gland of Decapoda. There are many Eumalacostraca such as the Isopoda and Stomatopoda which also have a maxillary gland in the adult stage. Both glands consist of a coelomic sac followed by an excretory duct, the nephridium, which generally is longer in fresh-water than in marine species There are secondary structures, such as the coxal and gill glands, as well as nephrocytes in several species.

Genital system and reproduction. The sexes are separate, and generally in the female the ovaries are elongate. Oviducts, receptaculum seminis, vulva, and vagina are present. The male system consists of the testis, vas deferens, ductus ejaculatorius, and sometimes a penis, modified thoracic limbs (pleopods), or petasma Spermatophores containing spermatozoa are of various form and size. There is a great variation in the secondary genital structures for copulation, brood protection, and finding of partners.

In *Lucifer*, *Euphausia*, and other Eumalacostraca, there is total cleavage and the formation of a hollow blastula. Other groups having a centrolecithal egg exhibit meroblastic cleavage with the formation of a superficial blastoderm. In telolecithal eggs, the protoplasm accumulates at one pole, forming a cellular disk (*see* OVUM). Primitive Eumalacostraca do not undergo metamorphosis or have only a naupliuslike stage. The free-living zoaea is the most characteristic larva of the Eumalacostraca. It gives rise to the metazoaea

The nervous system is typically arthropodian There is a great tendency to fusion of the ganglia. In the branchiuran Decapoda, 17 pairs of ganglia of the ventral nerve cord are united into one immense thoracic ganglion. The higher crustaceans also have a well-developed sympathetic nervous system. The main part is in the head region and consists of the ganglion viscerale and stomatogastricum, as well as the nervus cerebrovisceralis Eyes are of two types, the primitive median eye of the nauplius which occurs reduced in all the Eumalacostraca, and the pair of compound eyes which is very characteristic of higher crustaceans. These eyes are either sessile or stalked. They are sometimes lost during adaptation to a subterranean life as in Bathynellacea, some Asellidae, and Amphipoda. There is a great variety of tactile and chemical sense organs located mainly on the anterior part of body. Of extreme interest and taxonomic value are the static organs, or statocysts, for balance control.

Eumalacostraca are known for their great morphological variation which corresponds to their different modes of life. The majority of species are found in the sea, brine pools, and brackish water. In the large orders, adaptation of some species to fresh water is found. Some orders such as the Anaspidacea, Bathynellacea, and Stomatopoda are restricted to special biotopes. Of great interest are the species of Isopoda and Decapoda, which have become adapted to terrestrial life, as well as the parasitic Isopoda. Some marine forms are littoral; others comprise the marine plankton or are found in the abyssal region to depths of 4000 fathoms. All the chief groups are cosmopolitan except the Anaspidacea. Fossils are found in very old strata. *See* CRUSTACEA. [H.J.]

Bibliography: R. P. Bigelow, Report on the Crustaceae of the order Stomatopoda, *Proc U.S. Natl. Museum*, 17:489, 1891; S. J Holmes, Synopsis of the Amphipoda, *Am. Naturalist*, 37:267, 1903; H. Jakobi, Biologie, Entwicklungsgeschichte und Systematik von *Bathynella natans* Vejd , *Zool. Jahrb.*, 83:1–63, 1954; F. P. Paulmeier, *Higher Crustaceae of New York City*, N.Y. State Museum Bull. 91, 1905; M. J. Rathbun, *The Cancroid Crabs of America*, U.S. Natl. Museum Bull 152, 1931; H. Richardson, *Monograph of the Isopods of North America*, U.S. Natl. Museum Bull 54, 1905; G. Smith, On the Anaspidacea, living and fossils, *Quart J Microscop. Sci.*, vol. 53, 1909.

Eumetazoa

The phyla of the animal kingdom above the Porifera, or sponges. It includes those animals which have tissues or exhibit tissue formation and organ systems. As the Porifera are excluded, choanocytes, or collar cells, are not present in these phyla. A mouth and digestive tract are present in all except the parasitic species where these structures may be lacking. The Eumetazoa may be subdivided into the Radiata and Bilateria on the basis of symmetry. *See* ANIMAL KINGDOM. [C.B.C.]

Eumycetes

The true fungi, a subdivision or, in the opinion of some mycologists, a division of the plant kingdom. *Eumycetes are also known as the Eumycophyta and* are nonchlorophyllous thallus plants; that is, they lack roots, stems, and leaves. They are characterized by vegetative bodies or mycelia consisting of nucleated filaments, known as hyphae. Fungal mycelia are commonly termed molds. The Eumycetes comprise four classes:

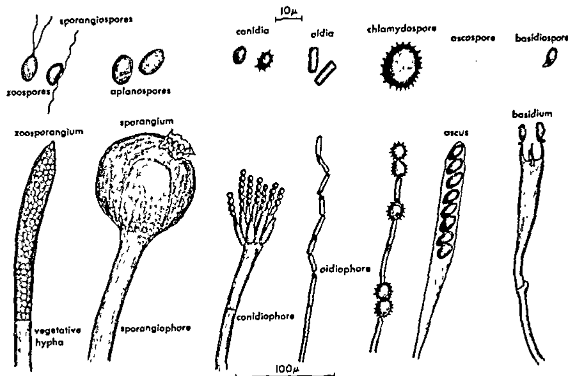


Fig. 1. Spore types in fungi.

Phycomycetes, class I, is the group in which the mycelium is made up of nonseptate hyphae. Examples are water molds, downy mildews, and black bread molds. *See* PHYCOMYCETES.

Ascomycetes, class II, is an order of fungi with ascospores, the final product of the sexual process. Examples are yeasts, powdery mildews, some blue and green molds, pink bread mold, cup fungi, morels, and truffles. *See* ASCOMYCETES.

Basidiomycetes, class III, is an order of fungi with basidiospores, the final product of the sexual process. Examples are rusts, smuts, bracket fungi, jelly fungi, mushrooms, puff-balls, earth stars, and stink horns. *See* BASIDIOMYCETES.

Fungi Imperfecti, class IV, is an order of vegetative molds, predominantly allied with class II, which lacks sexual reproduction. Examples are most blue and green molds, some yeasts, and most human pathogenic fungi. *See* FUNGI IMPERFECTI.

The Eumycetes and its four constituent classes can better be characterized, however, as an extensive array of combinations of specific expressions of several variable features. Variations of these basic features are outlined below with occurrence in the four classes indicated by corresponding class numbers.

Mycelium. The mycelium, generally the vegetative body of fungi, is extremely variable. Unicellular forms, thought to be primitive (I) or derived (II), grade into restricted mycelial forms; in most species, however, the mycelium is extensive and capable of indefinite growth. Some are typically perennial (III) though most are ephemeral. The mycelium may be nonseptate, that is, coenocytic (I), with myriad scattered nuclei lying in a common cytoplasm, or septate (II, III, IV), with each cell containing one to a very few nuclei (III) or an indefinite number of nuclei (II, IV). Septa may be either perforate or solid. Cell walls are composed largely of chitinlike materials except in one group of aquatic forms (I) that have cellulose walls. Most mycelia are white, but a wide variety of pigments can be synthesized by specific forms and may be secreted into the medium or deposited in cell walls and protoplasm. Mycelial consistency varies from loose, soft wefts of hyphae to compact, hardened masses that resemble leather. Each cell is usually able to regenerate the entire plant body, and vegetative propagation commonly results from mechanical fragmentation of the mycelium.

Reserve foods are stored in various fungi as fats or oils and glycogen (animal starch); volutin granules are the typical reserve material of yeasts (II, IV).

Asexual reproduction. Asexual reproduction, propagation by specialized elements that originate without sexual fusion, occurs in most species and is extremely diverse. The most common and important means of asexual reproduction are unicellular or multicellular spores of various types that swim, fall, blow, or are forcibly discharged from the parent mycelium. The more important types of spores are the following.

Sporangiospores are borne in unicellular sacs termed sporangia and are common in class I. Sporangia originate either by differentiation of vegetative cells or as more specialized, newly-formed structures, frequently at the end of an

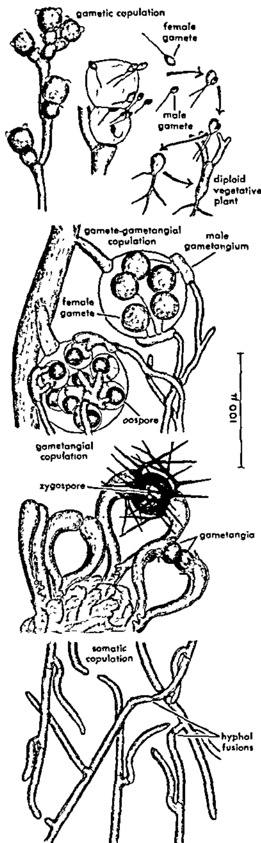


Fig. 2. Sexual mechanisms in fungi.

elongate stalk or sporangiophore. Sporangiospores are of two types: motile zoospores, equipped with flagella, and nonmotile aplanospores. The number of spores per sporangium varies from a few to thousands.

Conidia (II, IV) resemble nonmotile sporangiospores in shape, size, and structure but are produced externally upon a conidiophore which may be simple or quite elaborate.

Oidia (II, III, IV), small, thin-walled cells usually with flat ends, are produced by the autofragmentation of vegetative hyphae.

Chlamydospores (II, III) are thick-walled, nondeciduous spores interposed along vegetative hyphae.

Numerous other spore forms represent variants of the basic types above. Various pigments may be deposited in spore walls and account for mold colors in most cases. Asexual spores typically germinate by germ tubes that develop directly into vegetative mycelia.

Sexual reproduction. Sexual reproduction occurs in a majority of species of all classes except IV. Juxtaposition and fusion of compatible sexual cells are achieved by four distinct sexual mechanisms:

Gametic copulation (I) is the fusion in pairs of differentiated, uninucleate sexual cells or gametes formed in specialized sporangialike gametangia. In isogamy, the two members of the fusion pair are alike; in heterogamy, they are morphologically different.

Gametangial copulation (I, II) is the direct fusion of gametangia without the differentiation of gametes per se.

Gamete-gametangial copulation (I, II) is the fusion of a differentiated gamete of one sex with a gametangium of the other sex; the differentiated gamete may be either female (I) or male (II).

Somatic copulation (III) is the sexual fusion of undifferentiated vegetative cells.

Pattern of sexuality. This defines the underlying relationship between the two sexual elements that fuse in sexual reproduction; this relationship may be homothallism, which is fusion of elements of the same mycelium, or heterothallism, which is fusion only between elements of different mycelia. Obligatory cross-mating of heterothallic species is imposed by a number of distinct genetic devices that determine the mating competence of individuals in respect to sexual sign, that is, male and female (I, II), or to incompatibility class, that is, mating type (II, III). Incompatibility involves no morphological differentiation and in different groups may determine only two (II, III), scores (III), or even hundreds (III) of distinct self-sterile, but cross-fertile, mating types.

Product of sexual fusion. This and the time of nuclear fusion vary widely in the different classes and subgroups. Immediate nuclear fusion and the production of resting diploid spores, such as oospores and zygospores, or a diploid [illegible] phase commonly occur in class I and in [illegible] class II. In more advanced forms, [illegible]

leads to the establishment of nuclear pairs, each a dikaryon, in hyphal systems capable of limited growth (II) or indefinite growth (III). The dikaryotic mycelium of II is parasitic within the closed, flask-shaped, or cup-shaped fruiting body known as cleistothecium, perithecium, and apothecium, respectively, and its apical cells are differentiated as asci. The dikaryotic mycelium of class III is capable of independent, indefinite vegetative growth and comprises the fruiting body of mushrooms and puff-balls upon or within which the basidia are borne. Asci and basidia are saclike cells within which nuclear fusion and meiosis occur prior to production of endogenous ascospores and exogenous basidiospores, respectively.

Heterokaryosis. In heterokaryosis the association of genetically dissimilar nuclei in vegetative mycelia occurs in many—perhaps even in most—members of classes II, III, and IV.

Parasexuality. This is a vegetative process involving nuclear fusion and recombination of parental characters in the absence of sexual organs or fruiting structures; it has been demonstrated in classes II, III, and IV. [J.R.RA.]

Bibliography: E. A. Bessey, *Morphology and Taxonomy of Fungi*, 1950; D. H. Wenrich, I. F. Lewis, and J. R. Raper (eds.), *Sex in Microorganisms*, Am. Assoc. Advance. Sci., Publ. 37, 1955.

Euphausiacea

A well-defined order of the class Crustacea. These planktonic malacostracans are closely allied with, and sometimes included in, the Decapoda. The order consists of 2 families containing 11 genera with about 85 species. The euphausiids are predominantly pelagic, but some genera are neritic.

Although few in species, the euphausiids are large for planktonic animals, and, because of their size and very great numbers, they form perhaps the most important single element of the marine biomass.

The Euphausiacea are found in all oceans. However, they are sensitive to temperature, and in southern waters at least, are confined in recognizable groups according to the temperature of the water mass they inhabit.

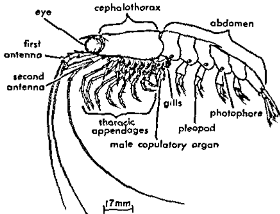


Diagram of euphausiid.

Known as krill to whalers, they constitute the diet of many of the whales, particularly the whalebone whales. The main whale feeding grounds coincide with the areas of greatest concentration of euphausiids, areas of convergence, backwaters, vortices of mixed layers, and centers of gyres. They are part of the diet of many other animals such as seals, herring, sardine, some birds, and even man. Their concentration in areas of high productivity is correlated with their own diet, which consists mainly of phytoplankton diatoms, microcrustaceans, and detrital matter. They are filter feeders, and their feeding mechanisms and mouthparts are elaborate.

Most euphausiids live at considerable depth during daylight, and many of them undertake extensive diurnal vertical migrations. The vertical position of these forms in the water column is thought to be photoregulated, and during the day they live under conditions of blue-green light of low intensity. Some euphausiids have been captured at depths greater than 2000 meters, and these bathypelagic forms have reduced eyes and attain considerable size, sometimes reaching 80 mm or more.

The euphausiids possess photophores which emit a brilliant blue-green light (*see* PHOTOPHORE GLAND).

The eyes are compound, bilobed in some genera, and contain three pigments, a carotenoid (astaxanthin), a melanoid pigment, and a photolabile pigment substance which is probably a visual pigment. The spectral sensitivity of the eye is greatest to blue-green light (*see* PHOTORECEPTION).

Respiration is by means of foliose, digitiform gills located at the bases of the second to eighth thoracic appendages.

The blood is a pale, leukocyte-bearing fluid with hemocyanin as the respiratory pigment. The heart is compact and has three pairs of ostia.

The male copulatory organs are extremely complicated and form the main criteria for specific identification. The phenomenon of swarming is common in many genera and is usually associated with reproduction. At copulation, the male transfers a spermatophore to the spermatheca of the female. The fertilized eggs are usually planktonic, although in some genera such as *Nyctiphanes* and *Nematoscelis*, the females retain their eggs in a type of brood pouch.

The larval stages are numerous and progress from the nauplius through two metanauplier, three calyptopis, about eight furcilia, and an undetermined number of cyrtopia stages. The animal may attain sexual maturity before the adult form is reached. Adult form is usually attained in one year, although some species, such as *Euphausia superba*, require two years. *See* MALACOSTRACA; SEA WATER FERTILITY. [B.P.B.]

Europe

Although long called a continent, in many physical ways Europe is but a great western peninsula of the Eurasian landmass. Its eastern limits are arbitrary

and are conventionally drawn along the water divide of the Ural Mountains, the Ural River, the Caspian Sea, and the Caucasus watershed to the Black Sea. On all other sides Europe is surrounded by salt water. The stretches of water between the Mediterranean and Black Seas, however, are narrow and it is possible to swim across the Bosporus to Asia Minor. Not even the Mediterranean Sea is a real physiographic boundary inasmuch as the Atlas Mountains of northwestern Africa are structurally part of the great Tethys geosyncline. Of the oceanic islands of Franz Josef Land, Spitsbergen (Svalbard), Iceland, and the Azores, only Iceland is regarded, mainly on historical grounds, as an integral part of Europe; thus the northwestern boundary is drawn along the Danish Strait. With these views in mind, the data on conventional limits and position of Europe are essentially those of Table 1.

Table 1. Conventional limits of Europe

Direction	Name and location	Latitude	Longitude
North	Nordkyn (mainland)	71°06′N	27°40′E
	Nordkapp (Magero Island)	71°12′N	25°45′E
South	Cap Tarifa (mainland)	36°00′N	5°36′W
	Cape Lithinos (Crete Island)	34°55′N	24°45′E
West	Capoda da Roca (mainland)	38°42′N	9°30′W
	Dunmore Head (island)	52°07′N	10°30′W
	Bjartangar (on Iceland)	65°30′N	24°28′W
East	Northern (Rocky) Ural	67°42′N	66°10′E

Maritime patterns. Europe is not only peninsular but has a large ratio of shoreline to land area reflecting a notable interfingering of land and sea. Excluding Iceland, the maximum north-south distance is 5680 kilometers (km); and the greatest east-west extent is 3860 km. Of Europe's area of 10,050,000 km^2 (that of United States with Alaska and Hawaii is 9,375,000 km^2), 73% is mainland, 19% peninsulas, and 8% islands. Altogether, Europe has a shoreline length of 37,900 km, that is, 3.55 km/1000 m^2. Also, 51% of the land is less than 250 km from shores and another 23% lies closer than 500 km. This situation is caused by the inland seas that enter, like arms of the ocean, deep into the northern and southern regions of Europe which thus becomes a peninsula of peninsulas. The most notable of these branching arms of salt water are the White Sea, the North Sea, the Baltic Sea with the Gulf of Bothnia, the English Channel (La Manche), the Mediterranean Sea with its secondary branches—the Tyrrhenian, Adriatic, and Aegean Seas—and finally, the Black Sea. Even the Caspian Sea, presently the largest salt-water lake of the world, formed part of the southern seas before the folding of the Caucasus. Of these inland seas only the Mediterranean and Black Seas have depths greater than 400 fathoms. The penetration of the landmass by these seas brings marine influences deep into the continent and provides Europe with a balanced climate favorable for the evolution and settlement of man. The favorable conditions are enhanced by the North Atlantic Drift (of Gulf Stream waters), which causes the northwestern coastal areas of Europe to be much warmer than the latitudes would suggest; for example, Labrador is much colder than the British Isles. (See Table 2.)

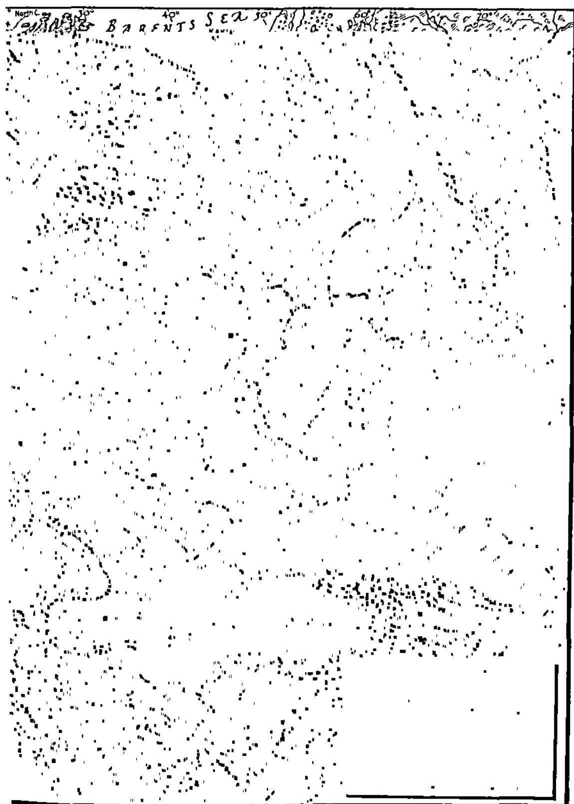
Land patterns. Europe has a unique diversity of land forms and natural resources. The relief, as varied as that of other continents, has an average elevation of 300 m as compared with North America's 440. However, 51% of the land lies below the 200-m contour lines, 27% is situated between 200 and 500 m, only 5% is higher than 1000, and only 1% exceeds 2000-m above mean sea level.

The shape and the over-all physiographic aspect of the great peninsula are controlled by geologic structure which delimits the major regional units (see Figs. 1 and 2). The triangular Central Lowlands (B on the map) with its acute apex in the Anglican Plains (B-15) and with younger strata covering the Precambrian Russian Table in the eastern parts (as in B-9, 10, 11) are bounded on each of the three sides by systems of mountain ranges with adjacent upland areas. On the northwestern side the erosional remnants of the once higher Caledonian ranges, folded during the early Paleozoic, are broken up into the Irish Uplands, Scottish Highlands and Uplands, and the Kjollen Mountains (A-3, 2, 1) bordering the predominantly granitic Baltic Shield of Precambrian age. In the east the late Paleozoic Variscian mountain revolution folded up the Ural Mountains (C-16) and welded them to the Russian Table. The southern side of the triangle is more complex. Here the Paleo-Mesozoic Tethys geosyncline was folded twice into mountain chains; the Variscian revolution produced the Hercynian Mountains and set the stage for the great Tertiary Alpine mountain revolution which elevated the chains of the Sierra Nevada, Pyrenees, Alps, Carpathians, Dinarides, Pindos, Balkan Ranges, and the Caucasus (E-27, 29, 30, 31, 35, 36, 32, and 33). These mountains form the Southern Highlands, whereas the eroded and block-faulted remnants of the Hercynian Mountains, from Iberia to the Lysa-Gora, complete the highly complex regional unit of Western Plateaus, Mountains, and Basins (E). (See Tables 3–7.)

Climate. This aspect of Europe is determined by a number of factors. Probably the most important are a favorable location between 35° and 71°N latitudes on the western or more maritime side of the world's largest continental mass; the west-to-east trend (rather than north-south) of the lofty southern ranges and the Central Lowlands, as well as of the inland seas which permit the prevailing westerly winds of these latitudes to carry marine influences deep into the continent; the beneficial influence of the North Atlantic Drift which makes possible ice-free coasts far within the Arctic Circle; and the low elevation of the northwestern mountain ranges and the Urals which allows the free shifting of air masses over their crests.



Fig. 1. Physical map of Europe. (Drawn by Erwin Raisz)


North C.
30°
40°
BARENTS SEA
50°
60°
70°

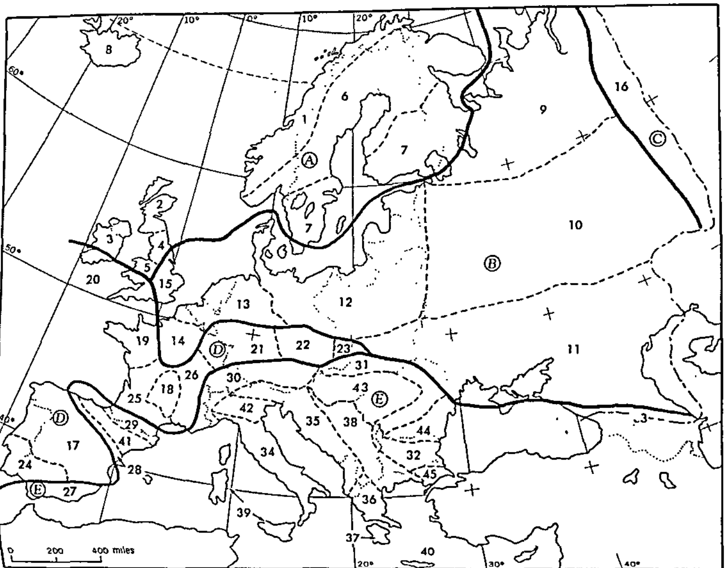


Fig. 2. Physiographic regions of Europe (*after G. Teleki*), showing major regional units and subdivisions. *A*, Northern Uplands: 1, Kjollen Mountains; 2, Scottish Uplands; 3, Irish Upland; 4, Pennines; 5, Wales; 6, Scandinavian Hill Land; 7, Southern Lake Region; 8, Iceland and Arctic Islands. *B*, Central Lowlands: 9, Arctic Lowlands; 10, Great Russian Plain; 11, Ukranian Steppes; 12, Baltic Lowlands; 13, North Sea Plains; 14, Paris Basin; 15, Anglican Plains C, Eastern Upland: 16, Ural Mountains. *D*, Western Plateaus, Mountains, and Basins: 17, Iberian Meseta; 18, Massif Central; 19, Armorican Upland; 20, Cornwall and South Ireland; 21, Ardennes, Shale Mts., Vosges, and Black Forest; 22, Bohemian Massif; 23, Lysa-Gora; 24, Portuguese Lowland and Andalusian Basin; 25, Aquitanian Basin; 26, Rhone Depression. *E*, Southern Highlands and Basins: 27, Baetic Cordillera; 28, Cantabrian Cordillera; 29, Pyrenees; 30, Alps; 31, Carpathians; 32, Balkan Ranges; 33, Caucasus and South Crimea, 34, Apennines; 35, Dinarides; 36, Pindos Mountains; 37, Peloponnesus; 38, Rhodope; 39, West Mediterranean Islands; 40, Aegean Islands; 41, Aragonian Basin; 42, Po Basin; 43, Carpathian Basin; 44, Wallachian Plain; 45, Thracian Basin.

Pressure, air mass, and moisture trends. The high atmospheric pressures built up in winter over central Asia obstruct the eastward passage of cyclonic storms, but the summer low of Asia, occurring simultaneously with the weakening of the Icelandic low, permits the prevailing westerlies to carry precipitation as far as the Urals and even beyond. Southeastward escape of these rain-bearing summer storms is barred by the Alps and Carpathians. Also, the seasonal shifts of air masses from and to the Icelandic and the subtropical Azores pressure centers produce precipitation maxima north of a line along the Pyrenees-Alps-Carpathian-Balkan Ranges and Caucasus mountains during the warm months of the year; whereas south of this line the summers are more or less rainless. These main climatic trends are broken up into a mosaic of local climates, depending on the closeness to shore lines, local elevation and slope direction, and influenced by human activities such as urbanization, industrialization, and agriculture.

Maritime climates. Among the marine climatic types, the Atlantic, the Mediterranean, and even the boreal polar climates are characteristic. True oceanic Atlantic climate occurs along the western fringes of Europe, from northern Portugal to central Norway and in the British Isles and Iceland. This west-coast marine climate is characterized by mild winters, with rarely a freezing period, and by somewhat cloudy, cool summers. The mean monthly range of temperature seldom exceeds 20°F. Precipitation, mostly in the form of drizzly rain, is regularly distributed over the year. Hence, runoff is small and soil erosion is at a minimum. Heather, wood anemone, small-leaved linden, English oak, and European beech are the representative plants of this climate They are also prevalent in central Europe and along some of the shores of the western Mediterranean. In the Iberian and the Italian peninsulas, as well as in southern France, this climate is tempered by warm Atlantic influences reaching as far north as southern England and Denmark and eastward into Greece and the Aegean archipelago, with a plant community characterized by the ivy, English holly, box tree, and fig.

Mitigated by oceanic climate along northwest Norway's irregular shores, the tundra climate of the Barents Sea coast occupies but a narrow coastal strip. Average temperature of the warmest month is below 50°F but above 32°F. Precipitation is scarce, the soil is permanently moist, and true summers are absent. Tundra vegetation consists

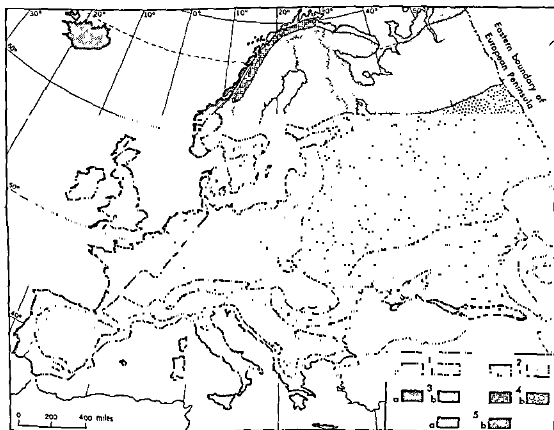


Fig. 3. Map of climatic regions of Europe. Regions and subdivisions are based upon combined weather and vegetation data 1. Boreal: a Polar, b. Subpolar; 2. Continental a. Forest and savanna, b. Steppe; 3. Marine: a. Atlantic, b. Mediterranean; 4. Mountain: a. High mountains, b. Plateau desert; 5. Transitional: a. Northern (wheat-rye), b. Southern (wheat-corn).

Table 2a. Seas, gulfs, bays, and fiords

Name	Main water body	Coast countries	Major rivers	Major harbors
Adriatic Sea	Mediterranean Sea East Basin	Italy, Yugoslavia, Albania	Po, Drin	Brindisi, Kotor, Rijeka, Trieste, Venice
Aegean Sea	Mediterranean Sea East Basin	Greece, Turkey	Vardar, Struma, Maritza	Pireus (Athens), Thessaloniki
Aegina, Gulf of	Aegean Sea	Greece		Pireus
Arcadia, Gulf of	Ionian Sea	Greece	Alfios	Kiparissia
Arctic Ocean		Norway, U.S.S.R.		
Atlantic Ocean		Iceland, Eire, U.K., *France, Spain, Portugal*	Tejo, Douro, *Garonne, Loire, Shannon*	Lisbon, Oporto, *Bordeaux, Nantes, Plymouth*, Reykjavik, Cardiff, Bristol
Azov, Sea of	Black Sea	U.S.S.R.	Don, Kuban	Kerch, Rostov
Baltic Sea		Denmark, Finland, Germany, Poland	Daugava (Dvina), Vistula, Oder	Gdansk, Leningrad, Stockholm, Riga, Szczecin, Copenhagen, Helsinki, Kaliningrad
Barents Sea	Arctic Ocean	U.S.S.R.	Pechora, Mezen	
Biscay, Bay of	Atlantic Ocean	France, Spain	Adour, Garonne, Loire	Bilbao, Bordeaux, St. Nazaire
Black Sea		Bulgaria, Romania, U.S.S.R.	Danube, Dniester, Bug, Dniepr	Constanta, Odessa, Sevastopol, Batumi
Bokn Fiord	North Sea	Norway		
Bothnia, Gulf of	Baltic Sea	Finland, Norway	Torneälu, Umeälu, Kemi	Vaasa
Bristol Channel	Atlantic Ocean	U.K. (England)	Severn	Cardiff, Bristol
Cadiz, Gulf of	Atlantic Ocean	Spain	Guadalquivir	
Caspian Sea		U.S.S.R.	Volga	Astrakhan
Cescaja Bay	Barents Sea	U.S.S.R.		Murmansk
Corinth, Gulf of	Ionian Sea	Greece		Korinthos
Drin, Gulf of	Adriatic Sea	Albania	Drin	
English Channel	Atlantic Ocean, North Sea	France, U.K. (England)	Seine, Somme	Calais, Dover, Le Havre, Portsmouth, Plymouth
Faxa Fiord	Atlantic Ocean	Iceland		Rejkjavik
Finland, Gulf of	Baltic Sea	Finland, U.S.S.R.	Narva (drains Peipus Lake)	Helsinki, Leningrad, Tallin
Firth of Forth	North Sea	U.K. (Scotland)		Leith
Firth of Lorne	Atlantic Ocean	U.K. (Scotland)	Loch Ness	
Gdynia, Gulf of	Baltic Sea	Poland	Vistula	Gdansk, Gdynia
Genova, Gulf of	Ligurian Sea	Italy		Genova (Genoa)
Greenland Sea	North Atlantic–Arctic Oceans	Norway (Spitzbergen, Jan Mayen)		
Irish Sea		U.K. (England), Eire		Dublin, Liverpool
Kandalaksa	White Sea	U.S.S.R.		
Kara Sea	Arctic Ocean	U.S.S.R.		
Kattegat	North Sea	Denmark, Sweden		Göteborg
Kiel, Gulf of	Baltic Sea	Germany		Kiel
Kurish Haff	Baltic Sea	U.S.S.R.	Memel	
Laconia, Gulf of	E. Mediterranean	Greece		
Ligurian Sea	W. Mediterranean	France, Italy	Arno	Genova, Livorno
Lions, Gulf of	W. Mediterranean	France	Rhone	Marseille, Toulon
Liverpool Bay	Irish Sea	U.K. (England)	Mersey	Liverpool
Marmara Sea		Turkey		Istanbul, Gallipoli
Mecklenburg, Gulf of	Baltic Sea	Germany		Lübeck
Mediterranean Sea		West Basin: Spain, France, Monaco, Italy East Basin: Italy, Yugoslavia, Albania, Greece, Turkey	Ebro, Segura, Rhone, Po	Marseille, Toulon, Gibraltar, Genova (Genoa), Pireus (Athens), Taronto, Trieste
Messina, Gulf of	E. Mediterranean	Greece		Messina
Moray Firth	North Sea	U.K. (Scotland)		
Naples, Gulf of	Tyrrhenian Sea	Italy		Napoli (Naples), Palermo
Nauplia, Gulf of	Aegean Sea	Greece		Nauplia
North Sea	Atlantic Ocean	U.K., Norway, Denmark, Germany, Netherlands, Belgium	Rhine, Elbe, Thames	Hamburg, The Hague, London, Bremerhafen

Table 2a. Seas, gulfs, bays, and fiords (Cont.)

Name	Main water body	Coast countries	Major rivers	Major harbors
Norwegian Sea	North Atlantic Ocean	Norway		Trondheim
Onega, Gulf of	White Sea	U.S.S.R.	Onega	
Oslo Fiord	Skagerrak (North Sea)	Norway	Glomma	Oslo
Patras, Gulf of	Ionian Sea	Greece	Achaeleos	Patras
Perekop, Gulf of	Black Sea	U.S.S.R.		
Pomorske, Bay of	Baltic Sea	Poland	Oder	Szczecin (Stettin)
Porsanger Fiord	Arctic Ocean	Norway		
Quarnero (gulf)	Adriatic Sea	Yugoslavia		Rijeka (Fiume)
Riga, Gulf of	Baltic Sea	U.S.S.R.	Daugava	Riga
Salerno, Gulf of	Tyrrhenian Sea	Italy		
Saloniki, Gulf of	Aegean Sea	Greece	Vardar	Thessaloniki (Salonika)
Saros, Gulf of	Aegean Sea	Turkey		
Setubal, Gulf of	Atlantic Ocean	Portugal	Setubal	Setubal
Skagerrak	North Sea	Denmark, Norway		Oslo
Taranto, Gulf of	Ionian Sea	Italy		Taranto
Trieste, Gulf of	Adriatic Sea	Italy, Yugoslavia		Trieste
Trondheim Fiord	Norwegian Sea	Norway		Trondheim
Tyrrhenian Sea	Mediterranean Sea, West Basin	France (Corsica), Italy	Tevere	Napoli (Naples), Palermo, Messina
Valencia, Gulf of	W. Mediterranean	Spain	Jucar	Valencia
Varanger Fiord	Arctic Ocean	Norway		
Väst Fiord	North Sea	Norway		
Venice, Gulf of	Adriatic Sea	Italy		Venezia (Venice)
White Sea	Barents Sea	U S.S.R.	Dvina	
Zuider Zee	North Sea	Netherlands		Amsterdam

Table 2b. Lakes

Name	Area, km²	Countries, area, km²	Elevation, m	Approx. depth	Important tributaries	Important cities
Balaton	660	Hungary	106	11	Zala (outlet, Sio)	Siofok
Berre, Étang de	156	France	13	11		
Biesbosch, Lake	200	Netherlands	0	?		
Bodensee, Lake of Constance	539	Germany 328 Switzerland 211	395	252	Rhine*	Konstanz, Bregenz
Byelosero, Lake	1,125	U.S.S.R.	122	10		
Caspian Sea	436,000	U.S.S.R., Iran	−26	946	Volga, Ural	Astrachan, Baku
Chiemsee	80	Germany	519	74	Alz*	
Como, Lake	145	Italy	198	410	Adda*	Como
Corrib, Lake	190	Eire	8	44		
Femunnen, Lake	205	Norway	910	?		
Fertö, Neusiedler, Lake	320	Austria 232 Hungary 88	113	4		
Garda, Lake of	370	Italy	65	346	(outlet, Mincio)	
Geneva (Leman) Lake	582	France 234 Switzerland 348	375	310	Rhone*	Geneva, Lausanne
Ilmen Lake	918	U.S.S.R.	18	3		
Inari, Lake	1,385	Finland	118	80	Ivalo	
Jalpug, Lake	220	U.S.S R.		?	Jalpug*	
Jannitza, Lake	100	Greece	410	22		
Ladoga, Lake	18,180	U.S.S.R. 10,166 Finland 8,014	5	250	Svirj* (from Lake Onega)	
Lugano, Lake of	49	Italy 18 Switzerland 31	274	279		Lugano
Luzern, Lake of		Switzerland	437			Luzerne
Lomond, Loch	71	Scotland		?		
Maggiore, Lake	212	Italy 190 Switzerland 22	194	372	Ticino*	Locarno
Mälaren, Lake	1,140	Sweden	1	64		
Mjösa, Lake	359	Norway	121	443		
Neagh, Lough	396	North Ireland (U.K.)	15	31		Antrim
Neuchâtel, Lake of	216	Switzerland	427	154		Neuchâtel
Ochrida, Lake	367	Albania 219 Yugoslavia 248	698	286	Drin*	

Table 2b. Lakes (Cont.)

Name	Area, km²	Countries, area, km²	Elevation, m	Approx. depth	Important tributaries	Important cities
Onega, Lake	9,550	U.S.S.R.	39	124		Petrozavodsk
Oulojärvi, Lake	1,002	Finland	123	?		
Päijänne, Lake	1,301	Finland	78	93		
Peipus, Lake	3,583	U.S.S.R	31	18	Narva*	
Prespa, Lake	286	Greece 100 Yugoslavia 186	906	54		
Ree, Lake	165	Eire	39	36		
Saimaa, Lake	1,200	Finland	76	58		
Sasik, Lake	195	U.S.S.R.	0	?		
Scutari (Shkoder), Lake	370	Albania 148 Yugoslavia 222	6	44	Zeta (outlet, Bojana)	Shkoder (Scutari)
Segosevo, Lake	1,200	U.S.S.R.	109	99		
Thingullatn, Lake	85	Iceland	625	?		
Trasimene, Lake	129	Italy	258	7		
Vänern, Lake	5,550	Sweden	44	98	Klar (outlet, Gotäelv)	Karlstad
Vattern, Lake	1,900	Sweden	88	119		Jönköping
Vierwaldstätter, Lake	114	Switzerland	437	214		
Zuider See (artificial)	?	Netherlands	0	?		Amsterdam
Zürich, Lake of	89	Switzerland	409	143		Zürich

* Tributaries which cross the lake.

Table 3. Peninsulas

Name	Area, km²	Countries situated on peninsula	Seas bounding peninsula
Apennine	149,000	Italy, San Marino, Vatican City	Adriatic, Ionian, Tyrrhenian
Balkan	468,000	Albania, Bulgaria, Greece, Romania, Yugoslavia	Adriatic, Ionian, Aegean
Brittany	23,700	France	Atlantic Ocean, La Manche
Calabria	15,000	Italy	Tyrrhenian, Ionian
Chalcidike	7,500	Greece, Mt. Athos Monastery	Aegean
Fosenhalvöen	5,400	Norway	Norwegian
Gallipoli	600	Turkey	Aegean, Marmara
Gibraltar	5	U.K., Spain	Atlantic Ocean, Mediterranean
Iberian, Pyrenean	584,000	Andorra, Portugal, Spain	Atlantic Ocean, Mediterranean
Istria	4,400	Yugoslavia	Adriatic
Jutland	23,800	Denmark	North, Baltic
Kanin (Canin)	10,500	U.S.S.R.	Barents
Kerch	3,000	U.S S R.	Black, Azov
Kola (Cola)	100,000	U.S.S.R.	Barents, White
Krim (Crimea)	25,300	U.S.S.R.	Black, Azov
Kurland	17,600	U.S.S.R.	Baltic
Peloponnesos	22,500	Greece	Ionian, Aegean, Gulf of Corinth
Salentine	5,000	Italy	Ionian, Gulf of Taranto
Scandinavian	800,000	Norway, Sweden	Baltic, North, Norwegian
Varangerhalvöen	5,900	Norway	Arctic, Barents

Table 4. Single islands of area greater than 1000 km²

Name	Name of island group	Area, km²	Political affiliation	Name of water body	Capital city, harbors
Corsica		8,722	France	Ligurian	Bastia
Crete		8,166	Greece	E. Mediterranean	Iraklion, Chania
Euboea (Evvoia)		3,812	Greece	Aegean Sea	Chalcis, Karistos
Fyn (Fünen)	Danish	2,976	Denmark	West Baltic Sea	Odense, Nyborg
Gotland		2,960	Sweden	Baltic Sea	Visby, Slite
Great Britain	British	224,240	U.K.	North Sea	London, Liverpool, Portsmouth, Edinburgh
Hinnöy	Vestraelen	2,198	Norway	Norwegian Sea	Harstad

Table 4. Single islands of area greater than 1000 km² (Cont.)

Name	Name of island group	Area, km²	Political affiliation	Name of water body	Capital city, harbors
Iceland		102,846	Iceland	Atlantic Ocean	Reykjavik, Akureyre
Ireland	British	82,459	U.K., Eire	Atlantic Ocean	Dublin, Belfast, Cobh
Kolguev		3,500	U.S.S.R.	Barents Sea	
Lesbos (Mytilene)		1,630	Greece	North Aegean Sea	Mitilini
Lolland (Laaland)	Danish	1,241	Denmark	West Baltic Sea	Nakskov, Gedser
Lomonossow Land (Aleksandra)	Franz Josef Land	19,700	U.S.S.R.	Arctic Ocean	
Majorca (Mallorca)	Balearic	3,411	Spain	West Mediterranean	Palma
Novaya Zemlya		91,800	U.S.S.R.	Arctic Ocean	
Öland		1,345	Sweden	Baltic Sea	Borgholm, Ottenby
Ösel (Saarema)		2,710	U.S.S.R.	East Baltic Sea	
Rhodes		1,404	Greece	East Mediterranean	Rodi, Lindo
Sardinia		24,090	Italy	Tyrrhenian Sea	Cagliari, Pausania
Senja (Senya)	Vestraelen	1,590	Norway	Norwegian Sea	
Sicily		25,740	Italy	Mediterranean	Palermo, Messina, Catania, Syracuse
Sjaelland (Zealand)	Danish	7,016	Denmark	West Baltic Sea	Copenhagen, Helsingör
Vaigach		3,700	U.S.S.R.	Barents Sea, Kara Sea	

Table 5. Major mountains: systems, divisions, and subdivisions

Names of mountains, divisions, subdivisions (group)	Highest peak, elevation, meters*
Albanian Alps	Maja e Hekuravet, 2600
Alps	Mont Blanc, 4807
Alps, Eastern	
Carnic Alps	Coglians, 2780
Gailthal, Alps of	Sandspitze, 2863
Dolomites	
Antelao group	Antelao, 3263
Marmolata group	Marmolata, 3342
Pale group	Cima di Vezzana, 3191
Julian Alps	Triglav, 2863
Karawanks	Grintovec, 2558
Noric Alps	
Gurkenthal, Alps of	Eisenhut, 2441
Pusterthal, Alps of	Coll' Alto, 3435
Tauern, Hohe (high)	Grossglockner, 3798
Tauern, Niedere (low)	Hochgolling, 2863
Tuxer Alps	Rast Kogel, 2760
Zillerthal, Alps of	Hochfeiler, 3510
Rhaetian Alps	
Albula group	Piz Platta, 2386
Bernina group	Piz Bernina, 4052
Oetzthal, Alps of	Wildspitze, 3774
Ortler group	Ortler, 3899
Raetikon	Scesaplana, 2967
Silvretta group	Fluchthorn, 3403
Stubaier Alps	Zuckerhutl, 3057
Alps, Prealps	
Austrian Alps	
Eisenerz, Alps of	Hochtor, 2392
Hochschwab group	Hochschwab, 2214
Totes Gebirge (Mts)	Priel, 2514
Wienerwald	Schopfl, 890
Bavarian Alps	
Algau Alps	Widderst, 2536
Karwendel Gebirge	Birkkar, 2956
Lechthal, Alps of	Parseier, 3038
Wetterstein Gebirge	Zugspitze, 2964
Dauphinean Alps	
Lans, Mts. of	Grand Vaymont, 2346
Lure, Mts of	Roc de l'Aigle, 1616
Massif du Pelvoux	Les Ecrins, 4103
Vercors, Mts of	Mt. d'Ambel 1703
Lombardian Alps	
Bergamo, Alps of	Pizzo Arera, 2512
Brescia, Alps of	Dosso Pedalta, 1951
Chamonix, Alps of	Adamello, 3554
Provence, Alps of	Benarde, 1943
Massiv des Maures	Mont des Anges, 779
St Baume, Chain of	(1154)
Salzburg, Alps of	Hochkönig, 2938
Dachstein Gebirge	Dachstein 2996
Kitzbühel, Alps of	Katzenkopf, 2539
Savoy, Alps of	
Grande Chartreuse	Chamenchaude, 2087
Reposoir, Chain of	Pointe Percée, 2752
Styrian Alps	
Glein Alpe	Speikkogel, 1989
Kor Alpe	(2144)
Rax Alpe	Rax, 2009
Sau Alpe	Zirbitzkogel, 2397
Wechsel Gebirge	Wechsel, 1744
Schwyz, Alps of (Alps of Four Cantons)	
Eastern	Pilatus, 2132
Western	Le Vanil Noir, 2395
Venetian Alps	
Bellunese Alps	Col Visentin, 1764
Carnic Prealps	Coma dei Preti, 2703
Julian Prealps	Plauris, 1959
Alps, Western	
Bernese Alps	
Eastern	Jungfrau, 4166
Western	Wildhorn, 3264
Cottian Alps	
Northern	Mt Tabor, 3177
Southern	Mt Viso, 3841
Glarnian Alps	Todi, 3623
Graian Alps	Gran Paradiso, 4061
Massif de Vanoise	Grand Casse, 3081
Lepontine Alps	Monte Leone, 3552
Adula group	Adula, 3406
Maritime Alps	
Ligurian Alps	Mt Argentea, 3299
Pennine Alps	Matterhorn, 4505
Mt Blanc group	Mt. Blanc, 4807
Monte Rosa group	Dufour, 4663
Apennines	Mt. Vettoreni, 2448
Abruzzian Apennines	
Gran Sasso	Mt. Corno, 2921
Ernici, Mts	(1202)
Sabini, Mts	Pellechia, 1368
Simbrumi, Mts	Mt. Viglio, 215

Table 5. Major mountains: systems, divisions, and subdivisions (Cont.)

[illegible]	Highest peak, elevation, meters*
Calabrian Apennines	Montalto, 1956
Ligurian Apennines	Mt Maggiorarea, 1803
Lucanian Apennines	
Cilento, Mts.	Cervati, 1899
Maddalena, Mts.	Volturno, 1836
Marche, Apennines of	
Sibillini, Mts	Mt Vettore, 2478
Umbrian Apennines	Monte Pennin, 1570
Neapolitan Apennines	Miletto, 2050
Toscano Emilian Apennines	Cimone, 2163
Apuan Apennines	Pisanino, 1946
Ardennes	Botrange, 692
Bakony Forest	Körishegy, 713
Balkan Ranges	Jumruktsal, 2371
Balkan, Central	Jumruktsal, 2371
Srednja Gora	Bogdan, 1572
Srnena Gora	Rozovec, 1321
Balkan, Little	Blgarka, 1097
Balkan, Western	Midzor, 2186
Bihor (Bihar)	Great Bihor, 1849
Gyalu Mts	Nagyhavas, 1827
Metallic Mts	Csaklyakö, 1349
Cambrian Mts.	Seafell, 978
Cantabrian Mts	Torre de Cerredo, 2678
Lozara Mts	Rubia, 2214
Picos d'Europa group	Torre de Cerredo, 2678
Carpathians, Eastern	Gerlachovka, 2663
Barot Mts	Gorgo, 1018
Bistrite Mts.	Budasul, 1864
Ciucului (Csik) Mts	Tarhavas, 1662
Colimanu (Kelemen) Mts	Petrosul, 2102
Giurgului (Gyergyo) Mts.	Poiana, 1777
Gutin	Gutin, 1447
Hargita	Hargita, 1801
Lapus Mts.	Cibles, 1842
Rodna Mts.	*Pietros, 2305*
Trei Scaune (Haromszek) Mts	Lakoca, 1777
Forested Carpathians	Koverla, 2058
Southern Carpathians	
Brasov Mts	*Bucec, 2506*
Carnei Mts	Semenic, 1449
Cibin Mts	*Pareng, 2529*
Fagaras Alps	*Negoi, 2544*
Hateg Mts	Peleaga, 2511
Sebesului, Mts	*Cindrel, 2245*
Western Carpathians	
Beskides, Eastern	*Rawka, 1303*
Beskides, Western	*Babiagora, 1725*
Bukk Mts	*Istalloskö, 959*
Cserhat Mts	Csóványos, 939
Fatra, Great	Fatra, 1591
Fatra, Little	Strasov, 1214
Little Carpathians	Vysoka, 754
Magura Mts	Banikov, 2178
Matra Mts.	*Kékes, 1010*
Slovakian Ore Mts	*Poljana, 1486*
Tatra, High	Gerlachovka, 2663
Tatra, Low	Giumbir, 2045
Vepor, Mts	Poljana, 1459
White Carpathians	Javorina, 968
Caucasus	
Western Caucasus	Elbrus, 5660
Eastern Caucasus	Kazbek, 5045
Cevennes (see Massif Central)	
Czech Forest	Arber, 1457
Dinarides	Durmitor, 2528
Dinara Planina	
Staretina Planina	*Cvrisnica, 2228*
Vran Planina	Cincer, 2006
Grmec Planina	Klekovaca, 1961
Javor Planina	Zep, 1537
Kapela Planina	Bittorai, 1985
Prenj Planina	(2102)
Trusina Planina	Bjelasica, 1867
Snjavina Planina	Durmitor, 2328
Velebit Planina	Vaganski Vrh, 1758
Vranica Planina	Vranica, 2107
Dolomites (see Alps)	
Frankenwald	Dobra, 794
Erei, Mts. of Sicily	Mt. Altesina, 1194
Erzgebirge (Czech Ore Mts)	Klinovec, 1244
Fichtelgebirge	Schneeberg, 1501
Gennargentu Mts. (Sardinia)	Punta la Marmora, 1837
Gibellini, Mts. (Sicily)	Mt. Grande, 785
Grampian Mts	Ben Nevis, 1343
Harz	Brocken, 1142
Istrandza Mts	Magiada, 1035
Iberian Mts	Sierra del Moncayo, 2315
Jura Mts	Cret de la Neige, 1728
Kopaonik Mts.	Milanov Vrh, 2140
Lysa Gora	Lysica, 611
Massif Central	
Dore, Mts	Puy de Sancy, 1886
Cevennes Vivarais, Mts. of	Mezenc, 1754
Forez, Mts of	(1640)
Margueride, Mts of	(1543)
Peloritani Mts. (Sicily)	Mt. Grande, 1374
Pindos Ranges	
Elicon (Helicon) Mts.	Parnassos, 2459
Tsumerka, Mts.	Tsumerka, 2393
Vardusia, Mts.	*Jiona*, 2512
Voion, Mts.	Zmolikos, 2637
Xerovuni, Mts.	(1471)
Pyrenees	*Pic d'Aneto, 3404*
Abodi, Mts.	d'Orby, 2017
Alberes, Mts	Nouloa, 1357
Central Pyrenees	
Maledetta group	Pico d'Aneto, 3404
Rhodope	Mus Alla, 2925
Dospat Dag	Sutke, 2188
Kursa Dag	(900)
Kustar Dag	(2177)
Osogovska Planina	Rujen, 2228
Pirin Planina	Jel Tepe, 2920
Rila Planina	Mus Alla, 2925
Tocagik Dag	Karlak, 1900
Voz Dag	(1300)
Schwarzwald (Black Forest)	Feldburg, 1493
Scandinavian Mts.	Glittertinol, 2552
Kjollen (Kiolen) Mts	Kebnekaisse, 2113
Borgefjeld	(1703)
Dovrefjeld	Snehättä, 2306
Hardangerfjeld	Hallingskarvet, 1933
Jötunheimen	Glittertind, 2552
Serra d'Estrela	Malhao, 1991
Sierra de Gata	Jañona, 1367
Sierra de Gredos	Plaza de Almanzor, 2661
Sierra Guadarrama	Peñelava, 2406
Sierra Morena	Almenara, 1798
Sierra Nevada	Mulhacén, 3481
Sudeten	
Jesenik	Praded, 1490
Riesengebirge	Schneekoppe, 1603
Taigetos	San Elia, 2409
Tatra (see Carpathians)	
Tauern (see Alps)	
Thuringerwald	Gr. Beer, 983
Transylvanian Alps (see Carpathians)	
Ural Mts.	
Forested Urals (South)	Jaman Tau, 1645
Metallic Urals (Central)	Cosvenskij Kamen, 1688
Rocky Urals (North)	Töll-pos-is, 1769
Vatna Jökull	Oraefa, 2118
Vérfes-Pilis Mts.	Pilis, 757
Vosges	Guebwiller, 1426
Yaila Mts. Crimea (Krim)	(1543)

* Figures given in parentheses are elevations for peaks the names of which are not available

only of herbs, dwarf plants, and the characteristic tree of boreal regions, the dwarf birch.

On a relatively narrow strip along Europe's *southern coasts the Mediterranean climate* prevails. It influences to a certain extent some inland *areas as far north as a line through the Pyrenees,* Alps, Carpathian Basin, Crimea, and Caucasus. Mild winters with sunny weather and occasional rains change, through a delightful temperate spring, into dry and rainless summers with high temperatures and bright blue skies. Some of the most characteristic plants of this climate are associated in the brushy macchia formation; others are mostly cultivated plants, such as the vine, maize (corn), chestnut, olive, and orange trees.

Continental climates. The counterparts to the marine climates are continental climates extending *into Europe,* across the Urals, from the great Asiatic landmass with its climatic belts running *more or less parallel to northern latitudes.* Easterly air masses can move easily across the low, broad-backed Ural Mountains, and their impact reaches westward over Europe's central lowlands, where it mingles, especially in the summer, with the oceanic climate pushing eastward on the wings of the westerly winds. The northernmost zone of this continental influence has a subpolar continental climate, the southern boundary of which runs close

Table 6. Selected peaks higher than 2800 meters

Name	Elevation, meters	Location by mountain ranges, divisions and subdivisions, or island
Adamello	3554	Prealps, Lombardian Alps, Alps of Chamonix
Adula	3406	Western Alps, Lepontine Alps, Adula group
Aneto, Pico d'	3404	Central Pyrenees, Maledetta group
Antelao	3263	Eastern Alps, Dolomites, Antelao group
Argentea	3299	Western Alps, Maritime Alps, Ligurian Alps
Bernina, Piz	4052	Eastern Alps, Rhaetian Alps, Bernina group
Birkkar	2956	Prealps, Bavarian Alps, Karwendel Mts
Blanc, Mont	4807	Western Alps, Pennine Alps, Mont Blanc group
Casse, Grand	3081	Western Alps, Graian Alps, Massif de Vanoise
Cima di Vezzana (see Vezzana)		
Coll' Alto	3435	Eastern Alps, Noric Alps, *Alps of Pusterthal*
Corno, Monte	2921	Apennines, Abruzzian Apennines, Gran Sasso
Cristallo, Monte	3216	Eastern Alps, Dolomites, Tofano group
Dachstein	2996	Eastern Alps, Salzkammergut
Dufour	4663	Western Alps, Pennine Alps, Monte Rosa group
Écrins, Barre des	4103	Prealps, Dauphinean Alps, Pelvoux group
Elbrus	5660	Western Caucasus
Etna*	3279	Island of Sicily
Finsteraarhorn	4257	Western Alps, Bernese Alps
Fluchthorn	3403	Eastern Alps, Rhaetian Alps, Silvretta group
Gran Paradiso	4061	Western Alps, Graian Alps
Grand Casse (see Casse)		
Grossglockner	3798	Eastern Alps, Noric Alps, Hohe (High) Tauern
Grossvenediger	3660	Eastern Alps, Hohe Tauern
Hochfeiler	3510	Eastern Alps, Noric Alps, Alps of Zillerthal
Hochgolling	2863	Eastern Alps, Noric Alps, Niedere (Low) Tauern
Hochkönig	2938	Prealps, Alps of Salzburg
Jel Tepe	2920	Rhodope, Pirin Planina
Jungfrau	4166	Western Alps, eastern Bernese Alps
Kazbek	5045	Eastern Caucasus
Leone, Monte	3552	Western Alps, Lepontine Alps
Marmolata	3342	Eastern Alps, Dolomites, Marmolata group
Matterhorn	4505	Western Alps, Pennine Alps
Midi, Pic du	2885	Central Pyrenees
Mont Blanc (see Blanc, Mont)		
Monte Cristallo (see Cristallo, Monte)		
Monte Leone (see Leone, Monte)		
Monte Rosa	4630	Western Alps, Pennine Alps, Monte Rosa group
Monte Viso (see Viso, Monte)		
Mulhacen	3481	Sierra Nevada
Mus Alla	2925	Rhodope, Rila Planina
Olympos	2918	Eastern Greek Mts., Chazia
Ortler	3899	Eastern Alps, Rhaetian Alps, *Ortler group*
Pic du Midi (see Midi, Pic du)		
Pic d' Aneto (see Aneto, Pic d')		
Paradiso, Gran	4061	Western Alps, Pennine Alps
Parseier	3038	Prealps, Bavarian Alps, Alps of Lechtal
Rosa, Monte	4630	Western Alps, Pennine Alps, Monte Rosa group
Sandspitze	2863	Eastern Alps, Carnic Alps, Gailthal group
Scesaplana	2967	Eastern Alps, Rhaetian Alps, Rhaetikon
Tabor Mount	3177	Western Alps, Cottian Alps, northern
Todi	3623	Western Alps, Glarnian Alps
Tofana	3243	Eastern Alps, Dolomites
Triglav	2863	Eastern Alps, Julian Alps
Vezzana, Cima di	3191	Eastern Alps, Dolomites, Pale group
Viso, Monte	3841	Western Alps, Cottian Alps, southern
Wildhorn	3264	Western Alps, Bernese Alps, western
Wildspitze	3774	Eastern Alps, Rhaetian Alps, Alps of Oetzthal
Zuckerhutl	3057	Eastern Alps, Rhaetian Alps, Stubaier Alps
Zugspitze	2964	Prealps, Bavarian [illegible]

* Island location.

to 60°N latitude. Brief summers are marked by temperatures above 50°F, the relatively meager precipitation covers the ground for almost one-half of the year in the form of snow, and late spring floods often occur as a result of the blocking effect of frozen estuaries of rivers flowing northward into Arctic waters. The vegetation consists mainly of coniferous forests known as taiga. Norway spruce and Scotch pine are the main species, intermixed northward with the characteristic boreal tree, the dwarf birch.

South of this zone, in the Great Russian Plains, the long cold-winter climate predominates. During three months of the winter period, temperatures drop below the freezing point and the snow is blown along land surfaces by icy continental winds. Temperatures averaging over 50°F and the maximum rainfall occur in the June-to-August growing season, but precipitation rarely exceeds 20 in. per year. Scotch pine on sandy soil and deciduous forests prevail in the northern parts of this zone, whereas oak forests alternate with the fertile cropland on chernozem soils to the south.

The southernmost continental climatic zone, covering roughly the Ukranian Steppes, which are a western extension of the middle-latitude steppe and desert regions of Asia, is characterized by the semiarid steppe climate. With less than 5 in. of rainfall per year in the area of the lower reach of the Volga River, these eastern parts are a desert. Westward the 15-20 in. of yearly precipitation, with the maxima during the growing season, and mean annual temperatures just below 50°F predominate in the natural grasslands. The most characteristic plant of these regions is the Stipa grass which also appears to the westward in the Wallachian Plain and in the central portion of the Carpathian Basin.

Patterns of local climates. Among the more pronounced local climates of Europe, several are of importance. An outstanding example is the semiarid climate of the Iberian peninsula, partly due to the plateau character of the meseta upland and partly to warm Atlantic climatic influences. The peninsula has much scrubby vegetation and rocky soil covered with low grass.

Mountain climate is more a group of local climates than a single climatic type. Depending upon the elevation above sea level, the latitude in which they occur, the proximity to seas or landmasses, and the general slope directions, these climates show extreme variability. With glacial influences in the Alps, cordilleran character in the Pyrenees and in the southern Carpathians, subtropical influences in the western Caucasus, and boreal imprint in the Kjöllen Mountains, they have not much in common. Only the occurrence of a fir line and a snow line, due to temperature decrease in the lofty sections of these mountain ranges, gives them a certain uniformity. Local winds in such areas have special names, such as the föhn of the Alps and the nemere of the eastern Carpathians.

These climatic types intermix and compete in a nearly triangular area. This is comparable to Europe's structurally defined shape, with the apexes in the southern part of the Gulf of Bothnia, the western Pyrenees, and the western portion of the Black Sea. The center of this area is approximately near Wien, Austria. The northern transition zone is more tempered, the southern zone shows extremes of temperature, and towards the east both transitional climatic regions are apt to have variable extremes of precipitation which quite often cause flood and drought years.

Table 7. Volcanoes

Name	Year of last eruption	Elevation, meters	Rank by altitude	Geographic region
Askja	1926	698	6	Iceland
Etna	1923	3279	1	Sicily
Hekla	1846	1557	3	Iceland
Katla	1918	1600	2	Iceland
Santorin	1928	584	7	Cyclades, Greek Aegean
Stromboli	1934	926	5	Lipari Islands
Vesuvius	1944	1190	4	Apennine Peninsula
Vulcano	1926	500	8	Lipari Islands

Drainage patterns. The intricate relief and the climates of Europe are well reflected in the drainage system. Extensive drainage basins with large slow-flowing rivers are developed only in the Central Lowlands, especially in the eastern part. Streams with the greatest discharge empty into the Black Sea and the North Sea, although Europe's longest river, the Volga, feeds the Caspian Sea. Second in dimension is the Danube, which crosses the Carpathian Basin and cuts its way twice through mountain ranges at the Gate of Bratislava and at the Iron Gate. The Rhine and Rhone are the two major Alpine rivers with headwater sources close to each other but feeding the North Sea and the Western Mediterranean Basin, respectively. Abundant precipitation throughout the year, as well as the permeable soils and the dense vegetation which temporarily store the water, provides the streams of Europe north of the Southern Highlands with ample water throughout the seasons. The combined effects of poor vegetation, rocky and desolate limestone karstlands, and slight annual precipitation result in intermittent flow of the rivers along the Mediterranean coast, especially on the eastern side of peninsulas. Only the Alpine rivers carry enough water, and if it were not for the Danube and Rhone, both originating in regions north of the Alps, the only major river of the Mediterranean basin would be the Po. The main water divide between northern oceanic and southern Mediterranean drainage runs in a general east-northeast direction from Gibraltar along the Sierra Nevada, Cantabrian

Table 8. Major rivers

Name	Length, km	Drainage area, km^2	Approximate mean annual discharge estuary, m^3/sec	Discharges into	Main tributaries*	Source region and regions crossed
Bug (Polish)	779	73,470		Vistula	(R) Narew	Podolian Upland, Ukranian and Baltic Lowlands
Buh (Ukrania)	750	73,280		Black Sea		Podolian Upland, Ukranian Steppes
Danube	2,890	808,578	6,436	Black Sea	(L) Vah, Tisza, Olt, Siret, Prut; (R) Inn, Drava, Sava, Morava	Black Forest, Bavarian Lowland, East Alpine Foreland, Carpathian Basin, Iron Gate, Wallachian Plain, Dobrudja, Moldavia boundary
Dnieper	2,150	527,000		Black Sea	(R) Pripyat, Beresina; (L) Seim	Great Russian Plain, Ukranian Steppes
Dniester	1,370	76,900	600	Black Sea		Eastern Carpathians, Bessarabia (Ukranian Steppes)
Don	1,860	429,800		Sea of Azov	(R) Donets	Great Russian Plain, Ukranian Steppes
Donets	1,078	?		Don		Ukranian Steppes
Duero	780	98,375		Atlantic Ocean	(R) Esla	Iberic Mts. (Cantabrian Cord.), Iberian Plateau
Dvina	1,780	362,300		White Sea	Upper reach called Vicegda	Timan Mts., Arctic Lowlands
Ebro	930	86,000		Mediterranean Sea	(L) Segre	Cantabrian Cordillera, Aragonian Basin
Elbe	1,154	147,744	712	North Sea	(L) Saale	Sudeten Mts., Bohemian Basin, North Sea Plains
Garonne	650	84,800		Atlantic Ocean	(R) Lot, Dordogne	Pyrenees, Aquitanian Basin
Guadalquivir	579	57,120		Atlantic Ocean	(L) Genil	Baetic Cordillera, Andalusian Basin
Guadiana	820	67,840		Atlantic Ocean		Sierra Morena, Iberian Plateau, Portuguese Lowland
Inn	510	25,700	800	Danube	(R) Salzach	Bernese Alps, Engadin, Bavaria Alps and Lowland
Kama	2,030	?		Volga	(L) Bjelaga; (R) Vjatka	Permian Hill Land, Western Ural foothills, Russian Plain
Kemijoki, Kemi	494	50,000		Gulf of Bothnia	(R) Ounas	Maan Hills, Northern Hill Land
Kuban	824	?		Black Sea		Caucasus, North Caucasian foothills
Loire	1,002	120,500		Atlantic Ocean	(L) Vienne	Plateau Central, South Paris Basin, Armorican Upland
Memel	936	98,100		Baltic Sea		Pripyat Marshes, Baltic Lowlands
Meuse	925	33,000		North Sea		Plateau of Langres, North Sea Plains
Mosel	545	28,200		Rhine		Vosges, Hunsrück
Oder	903	124,671		Baltic Sea	(L) Neisse; (R) Warta	Sudeten Mts., Baltic Lowland
Pechora	1,500	320,300		Barents Sea	(L) Ishma	Northern Ural, Arctic Lowlands
Po	672	74,970	140–1,530	Adriatic Sea	(L) Adda, Ticino	Cottian Alps, Po Basin

Table 8. Major rivers (Cont.)

Name	Length, km	Drainage area, km²	Approximate mean annual discharge estuary, m³/sec	Discharges *into*	Main tributaries*	Source region and regions crossed
Rhine	1,326	224,400		North Sea	(L) Mosel; (R) *Neckar*, Main	Lepontine Alps, Vosges and Black Forest boundary, North Sea Plains
Rhone	812	99,000		Gulf of Lions	(R) Saone; (L) Durance	Lepontine Alps, Lake Geneva, Rhone Depression
Sava	907	100,519	1,171	Danube	(R) Drina	Julian Alps, Carpathian Basin
Seine	776	77,800		English Channel (La Manche)	(R) Marne, Oise	Plateau of Langres, Paris Basin
Sévern	338	21,000	280	Bristol Channel		Cambrian Mts, Wales
Tejo, Tagus	1,010	80,950		Atlantic Ocean		Iberian Meseta, Portuguese Lowland
Thames (Themse)	336	15,310	280	North Sea		*Cotswold* Hills, Anglican Plains
Tiber, Tevere	405	17,169		Tyrrhenian Sea		Tuscan Apennines, Apennine foothills (East)
Tisza	961	157,135	820	Danube	(L) Szamos, Maros, Koros (two sources, Black and White Tisza)	North-Eastern Carpathians, Carpathian Basin
Ural	2,380	219,900		Caspian Sea	(L) Ilek	Southern Ural, Kirghis Steppes, Caspian Depression
Vardar	322	?		Aegean Sea	(R) Crna	Albanian Alps, Thessalonian Plain
Vistula	1,387	198,290		Gulf of Gdynia	(R) Bug	Beskides (Western Carpathians), Baltic Lowlands
Volga	3,570	1,420,000		Caspian Sea	(R) Oka; (L) Kama	Valdai Plateau, Great Russian Plain, Ukranian Steppes, Caspian Depression
Warta	760	54,400		Oder		Lysa Gora foothills, Baltic Lowlands

* (L) Left bank, (R) Right bank

Mountains, Pyrenees, Massif Central, Western Alps, Black Forest, Czech Forest, Northern Carpathians, Pripyat Marshes, and Valdar Hills to the northern central Urals. The Urals form on their flanks the water divide between Europe and Asia.

Summary of physical patterns. Geologic struc-

and their regions are delimited on the physical maps (Figs. 1 and 2) and in the sketch map of climate (Fig. 3). Every regional unit comprises regions with traits of similarity; but similarity does not mean uniformity. For example, the Alps have been sculptured by glaciers of the great Ice Age and their valleys are broad and long, whereas the Pyrenees and Caucasus, which bore few glaciers, have narrow, steep-sloped, V-shaped valleys which give them a rougher aspect than that of the much higher Alps. Each region has its own characteristics of local relief, vegetation, animal life, and mineral resources. Some of the minerals occur in Precambrian and Paleozoic rocks of the ancient shields, blocks, and mountain chains, for example, the iron ores of Scandinavia and of England, Alsace-Lorraine, Spain, and Bohemia, or the coal fields of England, the Ruhr, Silesia, and the Donets basin. Oilfields are mainly concentrated in the Carpathian region in Romania, Hungary, Poland, and Austria; a great variety of ores is bound to the metamorphic rocks and Tertiary volcanic areas affected by the Variscian and Alpine mountain revolutions; bauxite and manganese occur in pockets of the deeply weathered Mesozoic limestone areas, mainly in France, Hungary, and Yugoslavia;

and precious metals are most abundant in the Urals and the Bihar Mountains.

For the sake of brevity and reference accessibility Tables 2–9 present major characteristics of outstanding related physical features: Seas, gulfs, bays, and fiords (Table 2a); Lakes (Table 2b); Peninsulas (Table 3); Islands greater than 1000 km² (Table 4); Major mountains—systems, divisions, and subdivisions (Table 5); Mountain peaks higher than 2800 m (Table 6); Volcanoes (Table 7); Rivers (Table 8).

See AFRICA; ASIA; ATLANTIC OCEAN; BALTIC SEA; BLACK SEA; CONTINENT; EAST INDIES; MEDITERRANEAN SEA; TECTONIC PATTERNS. [G.T.]

Bibliography: J. Gottmann, *A Geography of Europe*, 2d ed., 1954; G. W. Hoffman (ed.), *Geography of Europe*, 1953; N. J. G. Pounds, *Europe and the Mediterranean*, 1953.

Europium

Element number 63, europium, Eu, is a metallic element belonging to the rare-earth groups. Its atomic weight is 152.00, and the stable isotopes, Eu^{151} 47.82% and Eu^{153} 52.18%, make up the naturally occurring element. Its existence was first suspected



in 1889, when Sir William Crookes found an absorption band in the spectra of samarskite. He attributed this band to a new element which he named "S." In 1896, the element was found by E. Demarçay and he named it europium after Europe. "S" and europium were later shown to be the same element. Its compounds were first prepared in the fairly pure state by G. Urbain and H. Lacombe in 1904. The trivalent salts have a very pale pink color. A divalent series of compounds is known. Europium is the rarest of the naturally occurring rare earths. During the period 1936–1941, H. N. McCoy published a series of papers giving methods for preparing divalent europium salts. He succeeded in separating several hundred grams of europium salts from rare-earth concentrates which he had accumulated over a number of years. He gave or loaned this material to a large number of research workers, and as a result the properties of europium are now better known than the properties of many other rare earths. The metal is the most volatile of the rare earths and has a considerable vapor pressure at its melting point. Reduction of europium salts with calcium or alkali metals usually produces divalent europium salts. The metal is best prepared by vacuum distillation of a mixture of europium oxide and lanthanum metal. The metal is very soft, it is rapidly attacked by air, and really belongs more to the strontium-calcium-barium series than to the rare earth series. It probably has only two electrons in its conduction bands instead of three. Europium has a very high cross section for neutrons and for this reason is of considerable interest to the atomic industry for control rods and nuclear poisons. *See* RARE-EARTH ELEMENTS. [F.H.SP.]

Eurotiaceae

A family of ascomycetous fungi of the order Eurotiales in which the asci or spores are borne in well-developed, sometimes hard, completely closed fruiting bodies known as cleistothecia. The family was previously known as *Aspergillaceae*. The family contains about a dozen genera, five of which, *Eurotium*, *Sartorya*, *Emericella*, *Carpenteles*, and *Talaromyces*, are genetically linked with *Aspergillus* and *Penicillium*. Chiefly because of this connection, the Eurotiaceae are of considerable importance. *Sartorya fumigata* (= *Aspergillus fumigatus*) causes aspergillosis in man and animals. Among the other genera in this family are *Lilliputia* with a *Gliocladium* conidial stage; *Monascus*, the pink mold which often invades silage; and *Magnusia*, with long, curled appendages on the ascocarp *See* EUROTIALES; FUNGI; MYCOLOGY, MEDICAL. [C.J.A.]

Eurotiales

An order of fungi of the class Ascomycetes bearing ascospores in globose or broadly oval, evanescent

The order is divided into the following six families:

1. Gymnoascaceae is a family of some 11 genera of fungi growing on dung, soil, and feathers. In *Byssochlamys*, no ascocarp is formed, the asci being borne in tufts on the mycelium. In the other genera, the ascocarp is represented by a more or less loose cottony mass of mycelial filaments around the asci.

2. Eurotiaceae is characterized by a firm-walled ascocarp. This family includes the genera to which the sexual stages of *Aspergillus* and *Penicillium* belong.

3. Onygenaceae comprises fungi that inhabit various animal substrata, such as hoofs and horns. The stalked ascocarp bears a spherical head which contains the asci.

4. Trichocomaceae is a small, little-known, chiefly tropical family with ascocarps from which a tuft of capillitial threads extrudes, releasing the ascospores after dissolution of the asci.

5. Terfeziaceae are underground, trufflelike fungi found in the Mediterranean region.

6. Elaphomycetaceae is a family of underground, saprophytic or mycorrhiza-forming fungi, characterized by ascocarps with thick, usually woody walls.

Some authors include the Chaetomiaceae and Ophiostomataceae in this order. *See* ASCOMYCETES; EUROTIACEAE. [C.J.A.]

Bibliography: E. A. Gäumann, *The Fungi*, 1952.

Euryalae

A suborder of Ophiurida in which the vertebrae articulate by means of hourglass-shaped surfaces, and the arms are able to coil upward or downward in the vertical plane. There is usually a leathery integument in which calcareous granules or platelets are imbedded. Most species are found in deep water, and often the arms are tightly coiled about the branches of black corals upon which Euryalae feed. Of the three families, the Gorgonocephalidae often have branched arms; the Asteronychidae have a large disk and slender arms, and the Asteroschematidae have a small disk and stout arms. *See* OPHIURAE; OPHIURIDA. [H.B.F.]

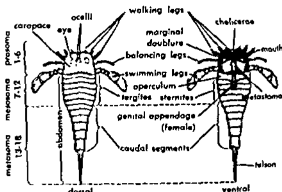
Euryapsida

A subclass of extinct reptiles of Permian and Triassic age formerly referred to as Synaptosauria. This group includes the well-defined order Sauropterygia (nothosaurs, plesiosaurs, and placodonts) and the poorly understood order Protorosauria. The skull of these reptiles is characterized by the possession of a single pair of upper temporal openings bounded laterally by the squamosal and postorbital bones. Other anatomical features are more difficult to appraise since the group includes delicately built terrestrial and moderately to highly specialized marine forms. The supposition that the sauropterygians had their ancestral roots among the protorosaurs is based on evidence that may well have to be reevaluated. *See* PROTOROSAURIA; SAUROPTERYGIA; *see also* REPTILIA. [R.Z.]

Eurypterida

Primitive aquatic arthropods of the subphylum Chelicerata and class Merostomata. The Eurypterids were dominant from the Ordovician to the end of the Paleozoic. Like other Merostomata their elongate-lanceolate bodies were encased in a chitinous exoskeleton comprising a cephalothorax (prosoma) and an abdomen (opisthosoma). *See* MEROSTOMATA.

The roughly cuboid cephalothorax has a pair of subcentral to anteriorly marginal compound eyes as the principal feature on the dorsal surface, in addition to the centrally situated ocelli. Ventrally, six pairs of jointed appendages are present, the first pair of which bears pincerlike chelicerae, whereas the last pair, in most genera, is developed as paddlelike swimming legs. The intervening four pairs are unspecialized walking legs in most genera, but may show various degrees of specialization. The bases of the appendages surround the centrally located mouth and are armed with masticatory organs. The chelicerae lie in front of the oral opening, which is bordered posteriorly by a small plate, or endostoma, while the larger metastoma overlies the postoral region.



Diagrams showing dorsal and ventral aspects of a typical eurypterid. (*Adapted from Clarke and Ruedemann, 1912; from R. R. Shrock and W. H. Twenhofel, Principles of Invertebrate Paleontology, McGraw-Hill, 2d ed., 1953*)

The abdomen may be pisciform or scorpioid in shape and is composed of twelve freely articulating segments and a terminal telson. The telson may be lanceolate, scutiform, stylioform, or a curved sting. In pisciform genera, body segmentation decreases gradually from front to back, but scorpioid forms show a sharp separation into a broad preabdomen and a narrow, cylindrical postabdomen.

On the ventral side, platelike appendages corresponding to the second through sixth abdominal segments bear gills, and the first of the series, the operculum, also carries the sexually dimorphic genital appendages, located medially.

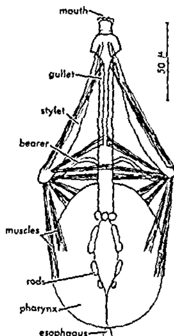
Most eurypterids were less than a foot in length, but several genera grew to giant size, approaching 6 feet. They inhabited fresh to brackish waters principally and most forms were bottom crawlers. However, certain streamlined genera (*Hughmilleria*, *Pterygotus*) were probably swimmers; others, actively raptorial, include those with powerful specialized prosomal appendages (*Pterygotus*), forms with possibly poisonous telson-sting (*Carcinosoma*), or genera with both (*Mixopterus*).

The Eurypterida, presumably descended from aglaspid ancestors, apparently left no direct descendants. [C.O.R.]

Eutardigrada

An order of tardigrades, lacking a cirrus lateralis, a sensory cephalic appendage, and clava, or club-shaped appendage. Pharyngeal pockets are strengthened by separated rods or macroplacoids, or are without thickenings. Claws are of different size, arranged in 2 pairs in which a larger and

smaller claw are united. In *Milnesium*, the claws are separated. *Haplomacrobiotus* has 2 simple claws. Eight longitudinal muscles are associated with the midgut. At the beginning of the hindgut, there are 3 excretory glands, called the vasa malpighii, which are specifically stretched or trilobed. Each gland is composed of 3 cells. The gonoducts open into the rectum, resulting in a single opening known as the anogenital pore or cloaca.



Buccal apparatus of Eutardigrada.

Thermozodium esakii, from a warm spring in Japan, evidently represents a third order, the Mesotardigrada Rahm (1937). This species combines such echiniscoidean features as the cirrus, clava, claws, and allusive plates with eutardigradan characters such as the pharyngeal rods, vasa malpighii, and cloaca. *See* HETEROTARDIGRADA; TARDIGRADA. [E.M.]

Eutheria

An infraclass of therian mammals including all living mammals except the monotremes and marsupials. Eutherians, often called placental mammals, are distinguished from their nearest relatives, the marsupials, by numerous characters. The full eutherian dentition, as seen in the mole for example, is

$$i\frac{3}{3}\quad c\frac{1}{1}\quad p\frac{4}{4}\quad m\frac{3}{3} = 44$$

The braincase is large, and the angular process of the mandible is not inflected. The young are born in a relatively advanced state of development, and are never sheltered in a pouch after birth. The Eutheria arose from pantotheres in the Cretaceous, but did not undergo their major evolution until the Tertiary. Living eutherians are divided into 16 orders; 10 additional orders are extinct and known only as fossils. The largest orders are the Rodentia, the Chiroptera, the Carnivora, and the Primates. *See* DENTITION; MAMMALIA; METATHERIA; THERIA. [D.D.D.]

Evaporation

The process by which a substance in the liquid state is converted into the vapor state. The molecules of substances in a condensed state are held to one another by strong forces of attraction, which are balanced by equally strong repulsive forces. Tending to overcome the potential energy of attraction is the escaping tendency of molecules, which arises from their kinetic energy. The kinetic energy, and therefore the escaping tendency of molecules, is a function of temperature. At each temperature, a certain fraction of the molecules possesses enough kinetic energy to overcome the forces of attraction of surrounding molecules and to escape from the surface of the liquid. If the process occurs at constant volume:

$$\frac{n_v}{n_l} = e^{-\Delta E/RT}$$

where n_v is the number of molecules per ml in the vapor, n_l is the number of molecules per ml in the liquid, ΔE is the difference in molar internal energy of the gas and liquid, R is the gas constant, and T is the absolute temperature. As the molecules which possess excess kinetic energy evaporate from the liquid, the average kinetic energy of the remaining molecules decreases, and the temperature drops. In order to maintain the temperature constant, heat must be furnished to the liquid. Ordinarily, evaporation occurs, not at constant volume, but at constant pressure. The quantity of energy required to evaporate 1 mole of liquid at constant pressure is called the molar latent heat of vaporization, ΔH, and is related to the internal energy by the first law of thermodynamics, $\Delta H = \Delta E + P\Delta V$, where $P\Delta V$ represents the work done by the vapor in expanding to a volume $\Delta V = V_{gas} - V_{liq}$ against the atmospheric pressure P. The molar volume of the liquid is ordinarily negligible by comparison with that of the vapor, and the gas obeys the ideal gas law to a first approximation ($PV = RT$). Hence, the latent heat of vaporization is given by $\Delta H = \Delta E + RT$. It is a function of temperature, and is normally measured calorimetrically at the normal boiling point. For nonassociated liquids, the latent heat of vaporization is given approximately by Trouton's rule, $\Delta H/T_b = 22$, where H has the units of calories, and T_b is the normal boiling point on the absolute temperature scale (°K).

The following factors affect the rate of evaporation of a liquid: the rate at which heat is supplied to the liquid to furnish the latent heat of [illegible]tion; the rate at which the liquid is [illegible]

to the surface molecules having sufficient kinetic energy to escape, and the rate at which the vapor above the liquid is changed to provide the best conditions for escape of molecules from the surface of the liquid.

For the conversion of a substance from the solid to the liquid state *see* SUBLIMATION. *See also* COOLING TOWER; EVAPORATOR; EVAPOTRANSPIRATION; HEAT EXCHANGER; HUMIDIFICATION; LIQUID; VAPOR PRESSURE [N.H.N]

Evaporator

A device used to vaporize part or all of the solvent from a solution. The valuable product is usually either a solid or a concentrated solution of the solute. If a solid, the heat required for evaporation of the solvent must have been supplied to a suspension of the solid in the solution, otherwise the device would be classed as a drier. The vaporized solvent may be made up of several volatile components, but if any separation of these components is effected, the device is properly classed as a still or distillation column. When the valuable product is the vaporized solvent, an evaporator is sometimes mislabeled a still, such as water still, and sometimes is properly labeled, such as boiler-feedwater evaporator. In the great majority of evaporator installations, water is the solvent that is removed.

Uses. Evaporators are used primarily in the chemical industry. Common salt is made by boiling a saturated brine in an evaporator. The salt precipitates as a solid in suspension in the brine. This slurry is pumped continuously to a filter in which the solids are recovered and the liquid portion re-

[illegible]

chemicals and soluble, organic wood constituents. The liquor is disposed of by concentrating it in evaporators to a strength at which the liquor will burn in a boiler. This avoids a water-pollution problem, recovers the inorganic chemicals for reuse, and provides sufficient heat energy not only to operate the evaporator but also to supply other needs of the mill as well. In the alkali industry, salt brine passed through a diaphragm-type electrolytic cell yields a dilute solution of salt and caustic soda. Evaporation to a strength of about 50% NaOH purifies the caustic by causing precipitation of practically all of the salt. It also simplifies shipment by bringing about a more than fivefold reduction in volume. Further evaporation of the caustic solution to a final temperature of about 700°F produces a practically anhydrous product that freezes at about 600°F and is shipped as a solid. Evaporators are also widely used in the food industry, usually as a means of reducing volume to permit easier storage and shipment.

Classification. The vaporization of solvent requires large amounts of heat. Provisions for transferring this heat to the solution constitute the largest element of evaporator cost and the principal means of distinguishing between types of evaporators. Practically all evaporators fall into one of the following categories:

1. Those heated by a flame that burns below the liquid surface, and in which the hot combustion gases are bubbled through the liquid.

2. Those in which the flame and combustion gases are separated from the boiling liquid by a metal wall, or heating surface.

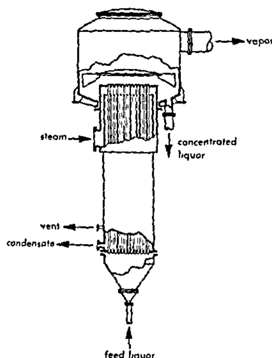
3. Those in which steam or other condensable vapor is the source of heat, and in which the steam condenses on one side of the heating surface and the heat is transmitted through the wall to the boiling liquid.

Submerged-combustion evaporators (type 1) are used primarily to concentrate solutions that would deposit a heat-insulating blanket of scale on the solid heating surfaces of other types of evaporators. They are also sometimes used for highly corrosive liquids. Submerged-combustion evaporators must use gas or oil as the fuel. Since the evolved vapor is mixed with the combustion gases, neither the solvent vapor nor its heat content can be recovered easily.

Direct-fired evaporators (type 2) are typified by the steam boiler and the old maple syrup kettle. They are not commonly used for the concentration of solutions, primarily because local overheating can cause formation of insulating deposits on the heating surface, which then becomes overheated and may melt or burn through. Also, a large heating surface is needed to recover the heat in the combustion gases and there is usually no cheap metal that will resist attack by both the combustion gases and the boiling liquid.

Steam-heated evaporators (type 3) are by far the most common, primarily because condensing steam gives up its heat so readily—condensing steam film coefficients are usually in excess of 1000 Btu/(hr)(ft^2)(°F). Thus the design of the evaporator and materials of construction can be suited to the solution being concentrated instead of being dictated by the problem of getting heat to the heating surface. The heating surface is usually in the form of metal tubes, since this represents the most economical method of putting the largest area in the smallest volume. The tubes may be vertical or horizontal, and the boiling liquid may be either inside or outside depending on the characteristics of the solution, such as viscosity, ratio of feed to evaporation, and whether or not a salt is deposited. *See* DISTILLATION.

Operation. The vaporization capacity of an evaporator is determined by the usual rules of heat transfer, and is directly proportional to the area of heating surface, to the difference between condensing steam and boiling liquid temperatures, and to the coefficient of heat transfer. The heat-transfer coefficient is usually limited by conditions on the boiling liquid side, although the condensing steam film and the resistance of the metal wall have some influence. Various means are employed to increase the boiling-film coefficient and all involve movement of the liquid relative to the heating



Long-tube vertical evaporator.

surface. A great many evaporators use only natural convection to accomplish this circulation. Typical is the long-tube vertical type shown in the illustration. Feed liquid enters the bottom of a nest of vertical tubes and begins to boil as it passes up the tubes. The boiling causes a large increase in volume, which accelerates the liquid to high velocities and gives good heat-transfer performance. The vapor-liquid mixture is separated in the chamber at the top of the tubes. The liquor may all be discharged as product, or part may be recirculated to the feed inlet.

A pump or agitator may be used to force the liquid past the heating surface if even higher heat-transfer coefficients are needed, for example, when corrosive conditions dictate use of the smallest possible area of an expensive alloy. Forced circulation evaporators are also used for scaling liquids or those from which a salt is to be crystallized, since there is less tendency for solids to form on the heating surfaces. Such evaporators usually consist of a flash chamber or crystallizing chamber, a conventional shell-and-tube heat exchanger, and a pump to circulate fluid from the chamber to the exchanger and back to the flash chamber. Another type of forced-circulation evaporator may be used to handle extremely viscous or foamy materials. This type either rotates the heating surfaces or employs a rotor that sweeps the materials around the walls of a stationary surface.

The water vapor evolved in an evaporator is usually about the same in quality and quantity as the steam used to heat the evaporator—the only difference being that the vapor has a lower pressure and hence lower condensing temperature. It is possible to compress the vapor so that it can be used as the heating steam in the same evaporator. Such thermocompression evaporators require far less energy for the compressor than would be needed to generate fresh steam. However, to keep the compressor cost and power consumption within reason, it is necessary to use a narrow compression ratio and this requires a large and expensive evaporator.

An alternative method of conserving heat is to use the water vapor evolved in one part of the evaporator to heat another part in which the liquid is boiled at a lower temperature. The water vapor evolved in this part, called an effect, can then be used to heat another effect boiling at still lower temperature, and so on until the boiling temperature is so low that the final vapor cannot be condensed by available cooling water. Such multiple effect evaporators are more expensive than single effect units because each effect can operate at only a fraction of the available total temperature difference. The total steam required, however, is roughly inversely proportional to the number of effects. The choice of the proper number of effects involves an economic balance of equipment costs and continuing steam costs. The great majority of evaporator installations employ the multiple-effect principle, and some have as many as 10 effects. *See* EVAPORATION; HEAT EXCHANGER; HEAT TRANSFER; SALINE WATER RECLAMATION. [F.C.SD.]

Bibliography: W. L. Badger and J. T. Banchero, *Introduction to Chemical Engineering*, 1955; J. H. Perry (ed.), *Chemical Engineers' Handbook*, 3d ed., 1950.

Evaporite (saline)

A sedimentary deposit of soluble salts resulting from the evaporation of a standing body of water. Quantitatively the most important evaporites are anhydrite, $CaSO_4$*; gypsum,* $CaSO_4 \cdot 2H_2O$*; and halite (rock salt), NaCl. Other evaporites, of much more limited distribution and volume but of economic significance, include potassium chlorides, sodium carbonates, borates, and nitrates.* *See* ANHYDRITE; GYPSUM; ROCK SALT.

Evaporites, being soluble, are rarely exposed at the surface except in arid regions. By far the greater part of the data on evaporites are derived from deep borings, chiefly those drilled in the search for oil and gas, and from underground workings developed to exploit economically valuable deposits. Gypsum, the mineral used in plaster and cement manufacture, is common near the surface but is invariably replaced by anhydrite at greater depths, the latter mineral being vastly more abundant. Other evaporite minerals are not related to depth of burial, except for the effects of near-surface solution.

Anhydrite and halite, the dominant evaporite minerals, occur in bedded deposits ranging from thin laminae to massive beds several tens of feet in thickness. These may be present in vertical succession, separated by partings of shale or carbonate to make up aggregate thicknesses of several

dreds, or even thousands, of feet. Such large accumulations are commonly interbedded with other strata bearing marine fossils which clearly demonstrate a relationship between major evaporite deposits and marine waters Typically, the evaporites are found in repeated cycles which approximate this vertical order from base upward: (1) marine fossil-bearing limestone; (2) dolomitized marine limestone; (3) fine-grained, finely laminated, unfossiliferous dolomite; (4) anhydrite; (5) halite. Many cycles lack the halite member, while in a few localities a sixth member bearing potash minerals is present.

Marine basin evaporites. The geographic distribution of major evaporite deposits and the thickness and character of the other rocks with which they are associated strongly indicate a relationship between evaporites and marine sedimentary basins which tended to subside during deposition of the sediments. It is estimated that more than 95 per cent of the volume of known evaporite deposits occur in such sedimentary basins. Two distinct patterns of basinal evaporite occurrence are noted: (1) basin-center evaporites occupying the interior of sedimentary basins; and, (2) basin-margin evaporites localized in a ring at the margin of the basins.

The basin-center occurrences commonly exhibit a concentric pattern in plan view, duplicating in a lateral fashion the vertical succession of the evaporite cycle, with marine carbonates (limestone or dolomite) at the outside and halite at the center. The peripheral carbonate belt in many instances includes elongate trends developed parallel to the margin of the basin and forming an encircling barrier. Such a barrier may be either of two types: (1) apparently wave-resistant trends of coral-algal reefs; (2) bar- or banklike trends of shell fragments, ooliths, and carbonate pellets which appear to have been built by wave action in shoaling water. Basins with extreme subsiding tendencies during evaporite deposition appear to have the greatest possibilities of containing potash salts. These are found with halite along the axis of maximum subsidence.

Basin-margin evaporites, usually anhydrite without large volumes of halite, occupy zones of varying width at the edge of a sedimentary basin. At greater distances from the basin these evaporites are typically intercalated with and pass into red shales and siltstones Toward the basin they tend to end relatively abruptly against trends of limestone or dolomite which may form an inner encirclement of the basin in the form of reefs, bars, or banks, as noted above. The interiors of such basins, within the encircling reefs, bars, or banks, are filled with marine limestones and shales, commonly dark in color and bearing a fossil fauna lacking the remains of bottom-dwelling organisms.

Theory of origin. Both major types of evaporite occurrence give evidence of forming in partially isolated bodies of water separated from the open sea by barriers established by organic growths (reefs) or by wave action (banks and bars) which restrict and confine the circulation of the waters. In the case of basin-center evaporites it is thought that evaporation of water in the basin produces an inflow of water from the surrounding or adjacent open sea. The inflowing water, of normal salinity and, therefore, of low density, flows readily as surface currents through passes in the barrier. Dense, highly saline water produced through evaporation sinks to the bottom and is unable to pass back through the barrier to mix with normal sea water (Fig. 1). Thus, sea water can continue to enter; but heavy brines are confined in the basin where they collect until the saturation points, first of the sulphates, then of the chlorides, are reached and precipitation of the salts takes place.

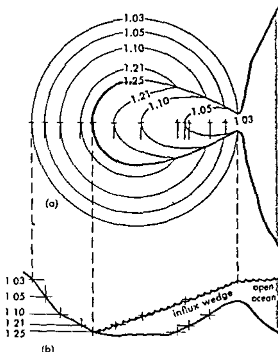


Fig. 1. Geometry of a model evaporite basin showing lines of equal brine density at a stage of saline equilibrium. (a) In plan view. (b) In cross section. *(From L. I. Briggs, Evaporite facies, J. Sediment. Petrol., 28(1)·46–56, 1958)*

Basin-margin evaporites appear to represent a reversal of this pattern of openly circulating and restricted waters. In the basin-margin case the waters of the basin interior are in open communication with the sea and are of normal salinity. At the edge of the basin, confined between barriers of reefs, banks, or bars, and a surrounding land area, marginal lagoons form evaporating pans. Here water passing through the barrier from the basin is confined, and evaporation leads to the precipitation of salts as noted above. It is apparent that basin-margin evaporites require an arid climate in the surrounding land area, since large streams of fresh water flowing into the lagoons would cancel the effects of evaporation.

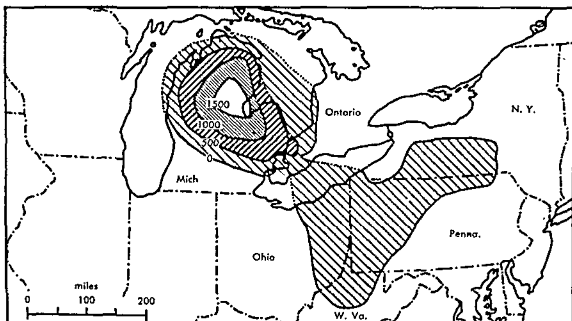


Fig. 2. Upper Silurian salt deposits as shown by drilling. The salt deposits are known to occur throughout most of the southern peninsula of Michigan, with thickness ranging upward to 1800 ft. Silurian salt beds aggregating 325 ft or less are distributed as indicated by the map. (*Data from USGS and from Fettke, Martens, Pepper, and Alling, Baltimore & Ohio Railway Co.*)

Shelf evaporites. Although basinal evaporites include by far the greatest volume of these salts, there are extensive deposits which show no relationship to subsiding sedimentary basins. The lateral continuity in thickness and character of relatively thin evaporite strata, and the other bedded sediments with which they occur, indicates deposition in or at the margin of shallow shelf seas covering broad areas of slow, uniform subsidence. Shelf evaporites, because of their wide distribution and the shallow depths at which they may be encountered, include many of the commercially exploited deposits of gypsum.

Two types of shelf evaporite occurrence are recognized; both are known as extensions of basin-margin deposits and also occur at great distances from basinal influences. One type is characterized by repeated thin cycles involving interbedded marine carbonates and evaporites. The second type is intercalated with red silts and shales which may contain remains of land animals and plants. The carbonate association is believed to represent deposition in shallow, ephemeral, evaporating pans at a distance from land areas or adjoining land areas of very low relief. The redbed type is interpreted as an alternation of land-derived detritus and the deposits of a shallow bordering shelf sea. *See* REDBED.

Nonmarine evaporites. Deposits suggesting precipitation from nonmarine waters are extremely rare in ancient rocks. However, present-day lakes, many of them temporary or seasonal, in areas of arid or semiarid climate are sites of evaporite deposition. Streams in such areas carry the soluble products of rock weathering to closed depressions occupied by ephemeral bodies of water where the salts are concentrated by evaporation. Nonmarine evaporites include locally important halite deposits and the major occurrences of borates, nitrates, and sodium carbonate.

Geologic distribution. In North America Cambrian evaporites appear to be confined to the arctic areas of Canada. Ordovician strata are notably poor in thick and extensive evaporite deposits. Strata of all subsequent geologic ages include major accumulations in one or more areas of North America. Certain sedimentary basins of long-continued subsiding tendencies contain thick evaporites representing several successive periods of geologic time (Fig. 2). Examples are the Michigan Basin (Silurian, Devonian, and Mississippian) and the Gulf Coast Basin (Permian ?, Jurassic, Cretaceous, and Tertiary).

[L.L.SL.]

Bibliography: W. C. Krumbein, Occurrence and lithologic associations of evaporites in the U.S., *J. Sediment. Petrol.*, 21(2):63–81, 1951; P. C. Scruton, Deposition of evaporites, *Bull. Am. Assoc. Petrol. Geologists*, 37(11):2498–2512, 1953; L. L. Sloss, The significance of evaporites, *J. Sediment. Petrol.*, 23(3):143–161, 1953.

Evapotranspiration

A term used to cover the total discharge of water by evaporation from land masses, including bodies of open water that form a part of the land masses. Applied to land areas as distinguished from water, it includes direct evaporation from

rock surfaces and transpiration by growing plants. It includes also perspiration by animals, but relatively this is very small. Of the precipitation received on the land, an estimated two-thirds is returned to the atmosphere by evapotranspiration.

Evaporation and transpiration are grouped together because of the difficulty of measuring them separately. Transpiration is the process by which moisture is drawn up through the roots of plants, circulated through the plant structure, and returned to the atmosphere in the form of water vapor. There are three main types of plants with respect to the water environment and transpiration: hydrophytes, mesophytes, and xerophytes.

Hydrophytes. Plants so designated live partly or wholly submerged in water, or with roots in saturated soil that is intermittently submerged These include cypress trees, water lilies, tules, cattails, and many others. Plants that habitually obtain their water supply either directly from the zone of saturation or from the capillary fringe extending upward from it are a subdivision of this group and are called phreatophytes Especially in parts of arid regions where the water table is shallow they are a prominent part of the plant life. They include, among many others, alfalfa, salt grass, mesquite, salt-cedar, willows, cottonwood, and greasewood They are virtually natural wells with pumping equipment capable of raising water from considerable depths Mesquite roots have been observed as much as 60 ft below the land surface, and alfalfa roots at more than 100 ft.

In the 17 western states of the United States it is estimated that phreatophytes alone transpire roughly 25 000,000 acre-feet of water annually (an acre-foot of water is about 325,400 gallons). In general, these plants use two to three times as much water as do useful mesophytes such as cotton and *wheat.*

Mesophytes and xerophytes. These varieties grow with their roots in well-aerated soil and draw their water supply from the moisture held in the soil by molecular forces. Mesophytes require a [illegible] to survive long dry periods without appreciable moisture. [A.N.S.]

Evergreen plants

Plants that retain their green foliage throughout the year. Popularly, needle-leaved trees (pine, fir, juniper, spruce) and certain broad-leaved shrubs (rhododendron, laurel) are called evergreens. In warm regions, many broad-leaved trees (magnolia, live oak) are evergreen, and in the tropics, most trees are evergreen and nearly all have broad leaves. Many herbaceous biennials and perennials have basal rosettes with leaves close to the ground that remain green throughout the winter. *See* FOREST AND FORESTRY; LEAF (BOTANY); PLANT TAXONOMY; *see also* BIENNIAL PLANTS; PERENNIAL PLANTS. [N.A.]

Evolution, organic

The theory that groups or kinds of organisms, such as species, may change with the passage of time so that descendants differ from their ancestors. According to this theory, modern plants and animals are the modified descendants of plants and animals that formerly existed, and these earlier organisms descended from still earlier and different forms. Only a relatively small number of forms, living in a given geologic period, left descendants, and of those that did so, some gave rise to more than one line of descent For this reason many different organisms among the descendants are regarded as related to each other in the sense that they have inherited from common ancestors. For example, all terrestrial vertebrates, the amphibians, reptiles, birds, and mammals, are believed to have descended, directly or indirectly, from a group of Devonian fishes (crossopterygians); at a later stage in earth history one group of reptiles (cotylosaurs) is believed to have given rise to evolutionary lines leading to various later reptiles, including dinosaurs, and to birds and mammals.

The theory of organic evolution contrasts with the idea that each kind of plant and animal was separately created and has remained in its original form; the latter idea is known as the theory of special creation.

EVIDENCES OF EVOLUTION

The evidences for evolution are found in homologous similarities, geographical distribution of plants and animals, fossils, and genetics.

Homologous similarities. If inheritance from common ancestry is a fact, many modern organisms should exhibit similarities inherited from those common ancestors. Fundamental similarities, not connected with similar habits or means of livelihood, are called homologous similarities.

Structural homology. The bones in the arm of man and the wing of a bird are fundamentally the same though the limbs are used for markedly different purposes (Fig. 1). Evidently the limb bones

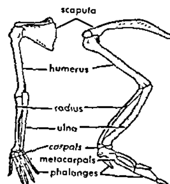


Fig 1. Homology of bones in the arm of man and the wing of a bird. Both limbs are drawn the same size for ease of comparison. (*From A. F. Shull, Evolution, 2d ed., McGraw-Hill, 1951*)

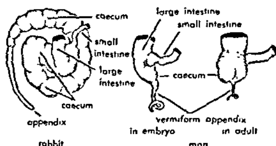


Fig. 2. Homology of cecum and vermiform appendix in rabbit, in human embryo, and in adult man. Drawn the same size for ease of comparison. (*After Bensley and Walter; from P. A. Moody, Introduction to Evolution, Harper, 1953*)

are modifications of a single pattern. This same pattern is found in the limb bones of all terrestrial vertebrates, despite modifications of the pattern connected with differing uses, such as climbing, running, flying, swimming, digging, grasping. Thus the limbs of all terrestrial vertebrates are homologous and their common pattern is regarded as evidence of inheritance from a common ancestor. This was the first vertebrate, descendant of crossopterygian fishes, to take up life on land.

Homology of vestigial structures. Homologous similarities involve all parts of the body and include both functional and useless or vestigial organs. Such a vestige is the human cecum and vermiform appendix, homologous to an extensive portion of the digestive system of various plant-eating mammals, such as rabbits (Fig. 2), whose cecum serves for storage of partially digested food while it is acted upon by bacteria. Vestiges seem most reasonably explained as inheritances from ancestors in which they were functional organs.

Embryonic homology. Homologous similarities are found in embryonic as well as in adult structure. Early embryos of all vertebrates resemble one another greatly; as development progresses the embryos become more and more unlike. The more closely related two vertebrates are, the longer are the portions of their embryonic developments which are similar. Furthermore, the embryos of higher animals pass through stages in which they resemble the embryos of certain lower animals. This phenomenon, called recapitulation, is regarded as evidence that the lower animals mentioned were ancestral to the higher ones. For example, at an early stage mammalian embryos, including human ones, resemble fish embryos in having, among other things, the structures from which gill slits form in fishes, a two-chambered (instead of four-chambered) heart, blood vessels (aortic arches) arranged *to supply gills (Fig. 3)*, and body musculature arranged in segments. These resemblances of the embryos are regarded as one evidence that remote ancestors of mammals were fishes.

Microscopic homology. Homologous similarities are found in microscopic structure as well as in larger features of anatomy. The fact that almost all plants and animals are composed of units called cells, and that cells in diverse organisms are similar in having nuclei with chromosomes and other common features, points to the fundamental unity of plan underlying all organisms. Cells divide by a complicated process of mitosis and form gametes by the process of meiosis. The steps of these processes are similar in all plants and animals.

Biochemical and physiological homology. The similarity between organisms extends to chemical structure. Cells are composed of the fundamental living substance, protoplasm, which is similar in chemical composition throughout the plant and animal kingdoms. Such complex substances as the high-energy phosphates and nucleic acids are nearly, if not altogether, universal. The chemical similarities are reflected in body fluids such as the blood. Blood tests reveal homologous similarities between the sera of animals, and are interpreted as indicative of inheritance from common ancestry. *See* BLOOD.

Homologous similarities are found in the processes, as well as in the structures, of organisms. The multitudinous chemical reactions involved in the processes of living, that is, metabolism, have great similarity among widely diverse organisms.

Geographical distribution. Evidence of evolution is also derived from the geographical distribution of plants and animals. Many forms are absent from regions which would be suitable for them, the absence being due to the fact that the region was not accessible to them. The unusual floras and faunas of oceanic islands clearly exhibit the role of

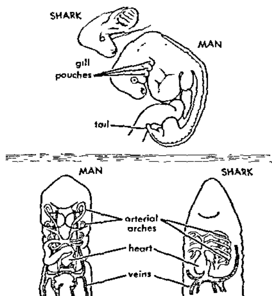


Fig. 3. Comparison of the embryos of fish and man, showing their similarity with regard to gill pouches and arterial (aortic) arches. Drawn the same size for ease of comparison. (*After A. F. Huettner in G. G. Simpson, C. S. Pittendrigh, and L. H. Tiffany, Life: Introduction to Biology, Harcourt, Brace, 19*

accessibility: only forms able to reach them from continents are there. Other forms for which the islands would be suitable are absent. Seemingly, a form arises in a certain region of the earth and then spreads to other suitable regions accessible to it. Frequently as the form spreads it changes somewhat, with the result that neighboring regions come to be populated by similar, but somewhat differing, varieties or species. Thus a series of forms may be found extending across one or more continents. At times some members of the series may become extinct, leaving related animals in widely separated territories, as, for example, the llamas in South America and the camels in Africa and Asia. If the record of fossils is complete the extinct forms will be known and the originally continuous distribution will be evident. Thus the fossils show that the camel family evolved in North America and spread to Asia, Africa, and South America. Camels then became extinct in North America, leaving a discontinuous distribution.

Fossils. The most direct evidence of evolution comes from a study of fossils, the remains of organisms that lived in past geologic ages (*see* PALEONTOLOGY) The geologic record reveals that in general later, more complex, and highly specialized forms were preceded by simpler, less highly specialized ones. Frequently a whole series of forms will be found extending throughout a section of geologic time Each member of the series differs but little from the forms immediately preceding and succeeding it, yet the total difference between the first member of the series and the last one is great. For example, *Eohippus*, the Eocene "dawn horse" (*Hyracotherium*), preceded *Mesohippus*, which in turn was succeeded by *Miohippus*, which was followed in succession by *Parahippus*, *Merychippus*, *Pliohippus*, and then *Equus*, the modern horse. Such a series is regarded as a record of evo- [illegible]

[illegible] In many cases the series of successive forms is not so fully known as it is in the evolution of the horse since the geologic record is far from complete. But as knowledge increases, "missing links" are found not to be missing [illegible]

[illegible] heritable variations in structure, breeders have developed a wide variety of cultivated plants and domestic animals. Many of the heritable variations utilized by breeders and experimenters arise as mutations, which are changes in the germ plasm that parents transmit to offspring (*see* MUTATION). Such mutations arise spontaneously at a low but fairly constant rate in both wild and cultivated organisms. They are generally considered to form the principal raw materials of evolutionary change.

At times mutations are combined with crossing, or hybridizing, of two varieties to produce new forms. G. D. Karpechenko crossed the cabbage with the radish, and eventually obtained a hybrid with double the number of chromosomes possessed by either parent; each parent has 18 chromosomes; the hybrid has 18 cabbage chromosomes plus 18 radish chromosomes. *See* BREEDING (PLANT); POLYPLOIDY. This hybrid is a true-breeding form with distinctive characteristics, and it is not interfertile with its parent species. It is called *Raphanobrassica* and constitutes the first new species deliberately produced by man.

MEANS AND METHODS OF EVOLUTIONARY CHANGE

Since the nineteenth century two main theories of the mechanisms of evolution have predominated

Lamarckism. In 1809 the French biologist J. B. Lamarck theorized that organisms are constantly striving to adapt to environmental conditions, which establish needs on the part of the organism. In striving to meet these needs they use certain parts and organs of the body, thereby causing these organs to become large, strong or otherwise well developed. Other parts and organs are not used and so become smaller, weaker, and less well developed. Such bodily changes are called acquired characters, since an organism achieves them by its own exertions in its attempt to adapt to its environment. Acquired characters, according to Lamarck, are then passed on to offspring of the organism that acquired them. The crucial point of the theory is the question of whether or not such individually acquired characters are in reality inherited by offspring. Carefully controlled experiments have failed to demonstrate that the effects of use and disuse of organs are inherited by offspring. Proponents of the theory maintain that what cannot be accomplished by man in a brief period can be achieved in a state of nature during the vast stretches of geologic time. Most biologists in the twentieth century have regarded the inheritance of acquired characters as not proven, though the theory has been revived in Soviet Russia under the leadership of T. Lysenko.

Darwinism. In 1858 Charles Darwin and Alfred Russel Wallace formulated the theory of evolution by natural selection, and in 1859 the theory was published by Darwin in *The Origin of Species by Means of Natural Selection*.

Natural selection is the theory that, in a state of nature, organisms are selected to survive and reproduce somewhat as a plant or animal breeder selects certain individuals in his stock to become parents of the next generation, and rejects other individuals for the purpose (artificial selection). The motive force of natural selection is thought to be the observed tendency of organisms to produce larger numbers of offspring than can survive under existing conditions. The tendency of populations to increase more rapidly than does the food supply upon which the populations depend was emphasized

by T. R. Malthus in an essay that influenced both Darwin and Wallace.

Under conditions of overpopulation in the face of limitations on food supply and other opportunities for livelihood a competition or "struggle for existence" will arise. This competition may be between members of the same species for food, territory, mates, and so on, or it may be between members of two different species for such things as food and territory. Also involved may be the striving to escape from predatory animals. In this struggle some individuals will survive and become the parents of the next generation, while others will fail. According to the theory the successful individuals will be predominantly those which happen to have characteristics that better fit them for the conditions of life than their competitors. Darwin stressed the point that organisms vary among themselves and that individuals happening to have favorable variations will on the average be the successful ones (see POPULATION GENETICS). If the offspring are to share the parents' superior characteristics, the variations must necessarily be heritable. Heritable variations are for the most part mutations. Hence natural selection is thought to operate upon mutations: individuals having mutations that improve chances for survival and reproduction are favored in the competition and so produce a disproportionately large share of the following generation. In this way all members of a population may, in some generations, come to possess a favorable mutation, and so be better adapted to their environment than were their ancestors.

If some of the organisms migrate into a different environment, or if the environment changes, a premium may be placed on characteristics different from the ones formerly favored. Natural selection operating in the new environment will favor different mutations, and thus in the course of time the organisms in the new environment will come to differ from those in the old. The differences may become sufficiently great to cause the form in the new environment to be regarded as a different race or species from the one in the old environment. In this way differing species may arise from common ancestors.

Geographic isolation. It will be noted that differences in environment are involved in the formation of new species. In most cases two populations that eventually become separate species, although they originally belonged to the same species, are probably separated from each other geographically (geographic isolation). Hence mutations that occur in one population are not shared with the other one and in time the differences between the populations increase.

Reproductive isolation. Eventually the populations may come to differ from each other so markedly that if and when members of one population come into contact with members of the other they will not interbreed. The two populations are then said to be reproductively isolated from each other. Reproductive isolation is an important hallmark of species and may take many forms, such as structural or physiological differences that render interbreeding impossible, differences in breeding season, differences in mating behavior (psychological or ethological differences), or production of sterile hybrids. Sometimes reproductive isolation may arise in the absence of marked structural differences, resulting in two populations that are much alike yet do not interbreed. The important point is that reproductive isolation keeps populations separate and distinct by preventing gene exchange between them. Such populations are considered separate species (*see* SPECIATION). According to the thinking of many biologists, by continuation and elaboration of the process described, new genera, families, orders and classes may also arise. *See* ANIMAL EVOLUTION; LIFE, ORIGIN OF; MACROEVOLUTION; PLANT EVOLUTION. [P.A.M.]

Bibliography: C. Darwin, *On the Origin of Species by Means of Natural Selection*, 1859; T. Dobzhansky, *Genetics and the Origin of Species*, 3d ed., 1951; T. Dobzhansky, *Evolution, Genetics, and Man*, 1955; P. A. Moody, *Introduction to Evolution*, 1953; G. G. Simpson, *The Major Features of Evolution*, 1953.

Exanthem subitum

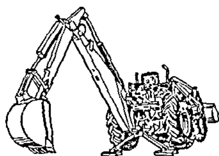
A mild, sometimes epidemic viral disease with abrupt onset, high fever, and rash. It is also known as roseola infantum. It occurs almost exclusively in infants 6 months to 3 years old. Sometimes exanthem subitum may be confused with German measles, from which it may be distinguished by its abrupt onset: distinction from other exanthematous diseases is chiefly by abrupt termination of fever before the rash appears. Older siblings and adults are thought to be either inapparently infected or already immune. *See* ANIMAL VIRUS; GERMAN MEASLES. [J.L.M.]

Excavator

Any of a variety of bucket-equipped construction machines used for digging earth and rock. The most common excavator is basically a modified crane (*see* CRANE HOIST). When modified for digging, the unit made up of crane and its digging attachment is commonly called by the name of the attachment. Thus a crane carrying a dragline bucket is called a dragline.

Standard excavators. There are four principal modifications of cranes for use as an excavator: dragline, clamshell, power shovel, and backhoe.

The dragline uses a long latticed crane boom, two working cables, and a bucket designed for horizontal digging. The bucket is cast into the area to be dug. The drag cable pulls the bucket toward the machine. Because of its weight and shape, the bucket fills itself. The filled bucket is lifted by the second cable and swung by the machine to the disposal area where release of the drag cable tips the bucket to empty it. The dragline is particularly suited for long-reach digging in wet and muddy areas where scrapers and bulldozers cannot work.



Hydraulic backhoe is used in trenching (Case)

The clamshell uses a crane boom and a clamshell bucket (*see* GRAB BUCKET). Digging with a clamshell is not a power operation; the clamshell is filled only by its own weight and the bite of its jaws as they are closed. For this reason, the clamshell is best suited for digging soft or loosely packed materials. It can, however, dig deeper and raise materials higher than any other crane-type excavator.

The power shovel uses a heavy but relatively

It is emptied by the opening of its hinged bottom. The power shovel is the most efficient excavator for hard or heavy digging within short reaches.

The backhoe is a short-boomed machine with an inverted bucket at the end of its stick. The bucket is filled by being pulled in an arc downward and backward toward the machine. The backhoe bucket is emptied by being raised and extended with the open side down. The backhoe is generally a smaller and lighter machine than the dragline or the power shovel; most backhoe buckets carry only fractions of 1 yd^3, as illustrated. The machine excels at digging trenches.

Special excavators. The hydraulic backhoe is a variation in which the operations are activated by hydraulic pistons rather than cables as illustrated. The basic machine may have either tracks or wheels and may be powered by either a gasoline or diesel engine.

The front-end loader consists of a continuous-track or wheeled tractor with a wide hydraulically operated bucket mounted on jointed arms at its front end. The bucket is filled by the forward motion of the tractor. It is raised and emptied by hydraulic pistons. The front-end loader is satisfactory for moving earth short distances or for loading trucks when long or high reaches are not involved.

The trencher is designed especially for excavating trenches. Basically it is a crawler tractor mounting either a movable wheel or a continuous chain, on which there are numerous buckets. The digging assembly is raised or lowered by cables to achieve the desired trench depth. As the assembly turns, the buckets scoop up earth, carry it upward, then spill it out onto a transverse conveyor belt which dumps it beside the trench. *See* BULK-HANDLING MACHINES; CONSTRUCTION EQUIPMENT.

[E.M.Y.]

Excitation

The application of voltage to an electric circuit or machine, an electron tube, or other device. This may be the input signal voltage itself in the case of a tube or amplifier. For motors, generators, electrodynamic loudspeakers, magnetic amplifiers, and similar devices requiring a constant or time-varying magnetic flux, the excitation voltage is the voltage that is applied to one or more coils to produce this flux.

Excitation of atoms or molecules of a gas involves the passage of an electron from one energy level to a higher level within the atom. *See* EXCITATION POTENTIAL.

Excitation of activity in living tissue or organs is achieved by applying electric pulses of controlled amplitude, duration, and spacing by means of appropriate electrodes. *See* BIOPOTENTIALS AND ELECTROPHYSIOLOGY. [J.MR.]

Excitation potential

The difference in potential between an excited atomic or molecular state and the ground state. The term is most generally used in connection with electron excitation. However, it applies equally well to the excited molecular vibrational and rotational states.

A closely related term is excitation energy. If the unit of potential be taken as the volt and the unit of energy as the electron volt, then the two are numerically equal (*see* ELECTRON VOLT). According to the Bohr theory, there is a relationship between the wavelength of the photon associated with the transition and the excitation energies of the two states. Thus the basic equation for the emission or absorption of energy is

$$\frac{hc}{\lambda} = E_i - E_f$$

where h is Planck's constant, c the velocity of light, λ the wavelength of the photon, and E_f and E_i the energies of the final and initial states, respectively. If the final state is the ground state, then the difference in energies is just the excitation energy of the initial state. If neither of the two states is the ground state, then the difference is numerically equal to the difference in excitation energies of the two states. This suggests that the excitation energy for many states may be determined spectroscopically. In fact, the careful measurement of the wavelengths associated with the transitions, along with the identification of the levels involved, permits the assignment of an excitation potential to each available energy state. This may be done by either emission or absorption spectra. *See* ATOMIC STRUCTURE AND SPECTRA.

Measurements have also been made using electron-impact means. Here electrons from a hot filament are accelerated by a grid and collected on an outer electrode. The potentials are so adjusted that the electrons can reach the collector only if they lose no energy after passing the grid. As the accelerating grid potential is increased, a series of

sharp drops is noted in the collector current. Each of these is interpreted as representing a case where the electrons have obtained just enough energy to produce excitation, and hence no longer have the energy needed to climb the potential barrier to the collector. The evidence thus obtained is very direct, but the accuracy is not as good as that for spectroscopic data. *See* GROUND STATE; IONIZATION POTENTIAL. [G.H.M.L.]

Bibliography: J. Millman and S. Seely, *Electronics*, 2d ed., 1951; F. K. Richtmyer and E. H. Kennard, *Introduction to Modern Physics*, 5th ed., 1955.

Excited state

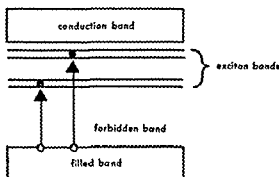
In quantum mechanics, a stationary state of higher energy than the lowest stationary state or ground state of a particle or a system of particles. Customarily, only bound stationary states, which generally are at most denumerably infinite in number, are spoken of as excited, although the formal quantum theory often treats the noncountable unbound stationary states on an equal footing with the bound states. Conventionally, the excited states are ranked in order of increasing energy; that is, the second excited state has higher energy than the first, which lies higher than the zeroth or ground state. Unlike the ground state, excited states frequently are degenerate. *See* DEGENERACY (QUANTUM STATES). *See also* GROUND STATE; METASTABLE STATE; STATIONARY STATE. [E.G.]

Exciton

An excited state of an insulator or semiconductor which allows energy to be transported without transport of electric charge. It can best be pictured using the band theory of solids. In this description crystalline materials which are nonmetallic have all their electrons in levels which cover a narrow band of energies. Each level in this band is filled with electrons, so that no electron can be excited to a higher energy level in the band. Above this filled band, or "valence band," is an energy region which in the case of pure materials has no levels and is called the "forbidden" band. Above the forbidden band is the "conduction" band which has energy levels in which electrons can move freely. In semiconductors and insulators the conduction band is normally empty. *See* BAND THEORY OF SOLIDS.

If a sufficiently energetic photon is absorbed by the crystal, an electron is raised from the filled band to the conduction band and is thus free to move. The empty level in the filled band is called a "hole" and has properties which are those of a free electron with a positive instead of a negative charge. Thus a free electron and a free hole are formed.

It is now known, however, that an electron and hole can be bound together and have a series of energy states in the same way that an electron and proton are bound in the hydrogen atom. An exciton is an electron and a hole in one of these bound configurations—usually the one of lowest energy.



Energy band diagram of an insulator. An exciton consists of the bound combination of an electron (black dot) that has been raised from the filled band, and the hole (white dot) that is left behind in the filled band because of the removal of the electron.

The computation of the energy levels for the exciton is similar to that for the hydrogen atom except that the hole substitutes for the mass of the proton, and the index of refraction of the solid enters the expression. The difference in energy in electron volts (ev) between the lowest exciton level and the bottom of the conduction band is given approximately by

$$E = 6.8/n^4$$

where n is the index of refraction. For sodium chloride the index of refraction is 1.5 and E is therefore about 1.3 ev. Germanium has an index of refraction of 4, leading to an E of 0.027 ev.

The exciton is free to wander through a crystal, just as a free electron does, and diffuses from place to place. A crude estimate of its velocity may be obtained by considering that it is in thermal equilibrium with the crystal lattice. Under these conditions

$$\tfrac{1}{2}mv^2 = \tfrac{3}{2}kT$$

where m is the mass of the exciton (which may be taken as two electron masses), v is the velocity of the exciton, k is Boltzmann's constant, and T is the temperature in degrees Kelvin. At room temperature the velocity of the exciton is approximately 10^7 cm/sec.

In low dielectric constant materials, such as the alkali halides, it is expected that the optical absorption band leading to the creation of excitons should be large. The first absorption band in the ultraviolet region of many of the alkali halides is indeed quite prominent and is generally attributed to exciton formation. Similarly, the decay of excitons by luminescence should be rapid and should lead to an exciton lifetime of about 10^{-8} sec. Arguments have been given which show that in high dielectric constant materials the absorption becomes rapidly weaker and the lifetime correspondingly longer.

The decay of excitons by luminescence is . . . tively rare, and the reason for this may be ... explained. Even in the alkali halides, which

a presumed exciton lifetime of 10^{-8} sec, the number of lattice sites passed by an exciton having a velocity of 10^7 cm/sec is nearly 10^7, since the distance between atoms is usually 1 or 2×10^{-8} cm. If any one of these 10^7 sites contains an imperfection, the exciton may be destroyed by transferring its energy to the imperfection. If this occurred, the exciton would ionize or emit luminescence characteristic of the imperfection. Therefore, only in very pure materials would one expect true exciton luminescence. Luminescence of this sort appears to have been observed in potassium iodide at liquid nitrogen temperatures (about 77°K).

In high dielectric constant materials the natural lifetime of the exciton is long and appreciable diffusion may occur. This appears to happen in cadmium sulfide, where diffusion distances of a centimeter have been observed under conditions which make it seem likely that the diffusing entity is the exciton.

It has been suggested that the transfer of energy along chains of organic molecules is also an exciton process and that this is an important element in the process of photosynthesis.

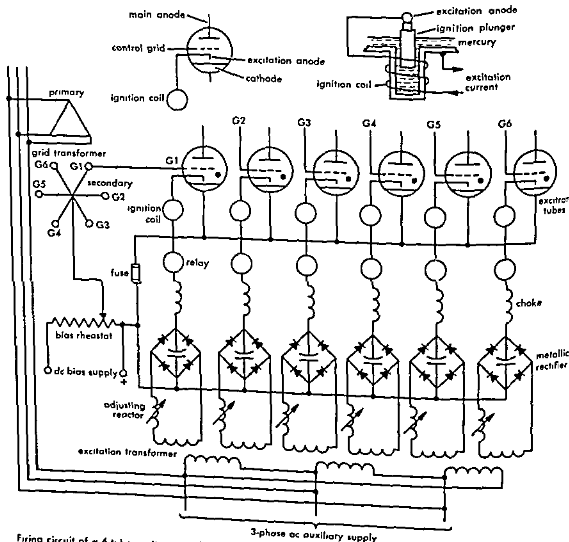
[C.C.K.; J.H.S.]

Bibliography: C. Kittel, *Introduction to Solid State Physics*, 2d ed., 1956.

Excitron

A single-anode pool-cathode tube provided with means for maintaining a continuous cathode spot. The arc starting and holding functions are similar to those used in multianode-tank mercury-pool rectifiers. *See* Mercury-vapor rectifier. Since the cathode spot is maintained continuously, means must be provided to keep it in the mercury pool and prevent its transfer to the metal wall. This may be done either by insulating the tank wall or by providing a cylindrical insulating shield supported from the cover.

The elements of a typical excitron excitation, or firing, circuit are shown in the illustration. The starter consists of a floating plunger, which makes contact with the excitation anode. When excitation



Firing circuit of a 6-tube excitron rectifier. (*Allis-Chalmers Manufacturing Company*)

current is flowing, the ignition coil causes the plunger to submerge and draw an arc from the excitation anode. The cathode spot on the plunger transfers to the mercury surface and remains there during normal operation. Phase control is obtained by varying the dc bias voltage in the grid circuit.

Excitrons have ratings and characteristics comparable to those of the ignitrons. *See* IGNITRON.

[C.C.H.]

Exclusion principle

This is frequently referred to as the Pauli exclusion principle after Wolfgang Pauli, who in 1925 first proposed that no two identical particles, such as two electrons, two protons, or two neutrons, can exist in the same quantum state at the same time. In an atom, for example, the state of any individual electron is characterized by four quantum numbers n, l, j, and m (or n, l, m_l, and m_s), and the exclusion principle requires that no two electrons have identical values of all four quantum numbers (*see* QUANTUM NUMBERS). As applied to electrons, the exclusion principle is basic to an understanding of the shell structure of atoms, of the energies of electrons in metals, of valency saturation in chemistry, and of the existence of positive electrons (positrons). As applied to protons and neutrons, it accounts for the shell structure of atomic nuclei and determines the condition of the interiors of stars.

In general, the state of any particle may be specified by giving its spatial wave function and the orientation of its intrinsic angular momentum, or spin. The elementary particles mentioned previously belong to a class in which the spin can have only two possible orientations with respect to a preferred direction, as for example the direction of a magnetic field. This limitation results from the fact that these particles have a spin angular momentum of $\frac{1}{2}(h/2\pi)$, and, as a consequence of space quantization, can set themelves only parallel or antiparallel to the magnetic field. Thus each possible spatial wave function represents two possible spin states. These correspond to the two allowed values of the quantum number m_s, namely $+\frac{1}{2}$ and $-\frac{1}{2}$. The remaining quantum numbers, n, l, and m_l, specify the spatial wave function. Because all electrons are identical, the magnitude of this wave function cannot be affected if any two of them are interchanged. This allows two possibilities; either the wave function preserves the same magnitude and sign after the permutation (symmetric wave function), or it keeps the same magnitude and changes its sign (antisymmetric function). According to the Pauli principle, if the spatial wave function is symmetric, it is impossible for the spin wave function also to be symmetric, since this would require the two electrons to be in exactly the same state. The total wave function of a state, including the spin, must therefore be antisymmetric. Particles having integral values of the spin, such as deuterons or light quanta. The exclusion principle is therefore equivalent to the statement that the complete wave function is antisymmetric, or that the particles involved obey Fermi-Dirac statistics. *See* BOSE-EINSTEIN STATISTICS; FERMI-DIRAC STATISTICS.

The importance of the exclusion principle lies in the fact that for systems to which it applies only a limited number of particles may be put in any state of given energy. Particularly for the states of lowest energy, the number is severely restricted. As a specific example, imagine electrons to be placed one by one in the neighborhood of a positively charged atomic nucleus. The first electron added can have only the one spatial wave function, specified by $n = 1$, $l = 0$, and $m_l = 0$. It may have m_s either $+\frac{1}{2}$ or $-\frac{1}{2}$. The next electron must take the opposite sign of m_s, and after it is added there are no further possibilities for $n = 1$ that satisfy the exclusion principle. The shell having $n = 1$ (the K shell in x-ray terminology) is said to be full. Additional electrons will have their lowest permitted energy for $n = 2$. Here l can be either 0 or 1, and after the 2 electrons having $l = 0$, $m_l = 0$, $m_s = \pm\frac{1}{2}$ have been added there is still room for 6 more, each having its m_l value either $+1$, 0, or -1, and one of the two possible values of m_s. The full complement for the L shell is therefore 8 electrons. Proceeding in this way, one can account for the shell structure of atoms, and for the periodic properties of the chemical elements.

The exclusion principle has other important consequences, one of the most striking being the behavior of the free electrons in metals. These electrons, being confined within the metal, have discrete energy states to which the Pauli principle applies. A great many of these possible states are completely filled, so that there is a relatively large zero-point energy, the maximum of which, called the Fermi energy, is several electron volts. For this reason, the free electrons do not contribute to the specific heat of a metal. *See* FREE-ELECTRON THEORY OF METALS.

[F.A.J.]

Bibliography: M. Born, *Atomic Physics*, 6th ed., 1957

Excretion

The removal of excess material from a cell or a living animal. Such a definition is so broad that it is necessary to set up limits to provide more useful descriptions and definitions.

Carbon dioxide is one of the chief products of the metabolism of cells. In unicellular forms the carbon dioxide diffuses away into the environment, and no special mechanism is present for its excretion. Even in the higher forms of life the carbon dioxide simply diffuses out of the cells into the tissue fluids and the blood, but special mechanisms have developed for its transport in blood. The excretion of carbon dioxide might properly be discussed here, but it has been found more satisfactory in the science of physiology t[illegible]

bon dioxide and oxygen together under the heading of blood-gas transport. This case is cited to exemplify the assignment of the excretion of heat to the subject of body-temperature regulation, of salts to osmotic regulation, and of urine and feces to the descriptions of the gastrointestinal and genito-urinary systems.

WATER

One of the central problems of excretory physiology is water. To understand the role of water, it is important to speculate briefly about the very origins of life. Current views assume life to have begun in the form of self-reproducing molecules which grew by incorporating the already-formed smaller molecules from the surrounding medium into their ordered structure. But, as essential ingredients in this medium diminished in quantity, self-reproduction became dependent upon a supply of the necessary material, obtained perhaps by modification of a closely related molecule. A primitive metabolism can thus be considered to have evolved. It would become essential to slow the diffusion processes which would inevitably carry away so valuable a material. This could be accomplished by the development of a relatively impermeable membrane from the many molecules which would be adsorbed onto the surface of a complex molecule of this sort. But when a semipermeable membrane is placed around a quantity of organic matter, osmotic pressure develops, and the primitive cell faces its first problem of excretion, namely that of getting rid of the water which osmosis continuously brings into it. One answer to this problem was that developed by plant cells. By surrounding the protoplasm with an extremely strong wall of cellulose, osmotic pressures of many atmospheres could be allowed to develop. The cell walls prevent the cell rupture that would otherwise follow. Whole plants take advantage of differences in osmotic pressure between external cells and internal cells to move water along the roots and stems. The turgor of cells, petioles, and leaves which develops as a result of osmotic pressure is also used to close stomata or to alter the position of a leaf. The development of cellulose cell walls has therefore been made useful in manifold ways. *See* OSMOREGULATORY MECHANISMS; PLANT PHYSIOLOGY.

Cellular forms which live in a salty environment may adjust their internal osmotic pressure to equal that of the medium, not by moving water, but by moving enough particles of various of the salts out through the cell membrane. All other equilibria may then be based upon diffusion. But it is clear that animal cells living in fresh water have no alternative except to keep pumping out the water that must pour into the protoplasm because of osmosis. Much of this pumping goes on at a level which has escaped measurement or even detection.

In some of the plant gametes and animals commonly placed at the bottom of the evolutionary array, protozoa and sponges, there is a conspicuous mechanism for ridding the body of the excess osmotic water. These structures are called contractile vacuoles. Measurements show that some animals may have to put out a volume of water equal to the total body volume every 4 minutes. The current view is that the water is secreted into channels in the protoplasm through which it is gathered into the vacuole for extrusion.

It is possible that the contractile vacuole represents simply the interiorization of a phenomenon common to most cells. The reverse process, the taking up of water from a medium, has been watched and called pinocytosis or cell-drinking. In such a case the surface membrane is very active. Drops of external fluid are trapped in infoldings of the membrane which then leave the boundary of the cell to become vacuoles in the protoplasm. If the membrane is digested away, the droplet becomes a part of the protoplasm. It has been suggested that the secretion of water may be accomplished by a reverse of this process. The many internal membranes that are now being revealed by the electron microscope surround droplets of water separated from the protoplasm. These vacuoles approach the surface membrane, incorporate into it, and discharge the droplet to the exterior. In the case of the contractile vacuole, the droplets discharge, instead, into the internal channels leading to the contractile vacuole because in these cellular forms it is important for the external pellicle to have considerable mechanical strength, whereas in the internal cells of multicellular animals the cell membrane can be a delicate and mobile one.

NEPHRIDIA

A most significant development in the evolution of excretory processes occurs in the phyla beyond the Protozoa, Porifera, and Coelenterata. Structures called nephridia are present in Platyhelminthes, Rotifera, Nemertea, Acanthocephala, Priapuloidea, Entoprocta, Gastrotricha, Kinorhyncha, Cephalochorda, and some Archiannelida and Polychaeta, as well as in larval stages of Polychaeta, Archiannelida, Echiuroidea, Mollusca, Phoronidea, and Cephalochorda. An eminent authority, E. Goodrich (1945), said that these organs (protonephridia) are in fact widely distributed and may be inferred to have been present in the common ancestor of all the metazoan Triploblastica.

Aside from the fact of its presence in all the animals at this intermediate level of evolution, the nephridium represents for the process of excretion the specialization of tissues that enabled animals to become larger and more independent of their immediate environment. The cells of the body no longer live in antagonism to sea water or fresh water, but are bathed by an internal medium which is regulated to represent an optimal environment. Numerous tissues have as their special function the regulation of the composition of this medium. It is the part of the nephridium to clear the body fluids of wastes, without at the same time losing from the body the substances needed by the cells. As in the case of the contractile vacuole, it is not understood

clearly how the nephridium accomplishes these functions, but assumptions can be made about several of the processes involved. See URINARY SYSTEM.

Two particular sorts of nephridia are worth describing briefly, the protonephridium and the metanephridium. The protonephridium is the primitive structure consisting of a single cell, shaped like a flask, within the neck of which there is found a cylindrical cavity or lumen. In this lumen are found one or more threadlike structures, or cilia, which beat continuously, the appearance thus giving rise to the old term, flame cell. The flame cell appears to filter through itself a part of the animal's circulating fluid, but its polarity is different from that of the other cells of the body and a "urine" passes from the lumen to the exterior via a nephridiopore.

The metanephridium is a structure for which the functions can be more nearly assigned. The tubular structure opens into a body or coelomic cavity There is always fluid in this cavity, resulting from filtration through the cells making up its lining. This fluid contains not only wastes but also materials of value to the animal, and as it is propelled down the nephridium by a lining of cilia it is subjected to changes as it passes the cells making up the tubule. Some salts and organic substances are reabsorbed before they leave the body. It is still not clear whether the tubule cells have the capacity to secrete materials into the incipient urine as it passes by; but it would be most surprising if they were not able to do so since other body cells, known as athrocytes, have been shown to have the capacity to accumulate high concentrations of dyes and metals.

HIGHER INVERTEBRATE KIDNEYS

Phylogenetically, nephridia are characteristic of primitive phyla and the larvae of some of the higher invertebrate phyla, but in the adults of all the higher phyla, except cephalochordates, they have been replaced by structures called renal organs or kidneys. The kidneys of the various groups are related to each other by their function, not by their structure or embryological relationships They are analogous rather than homologous organs.

Three major phyla are considered in the following sections—the Mollusca, Arthropoda, and Echinodermata.

Mollusca. It may be recalled that the trochophore larvae of the mollusks possess nephridia. As the larvae settle down and undergo metamorphosis the nephridia are lost and replaced by kidneys proper. It will also be recalled that a filtrate through the coelomic lining begins the process of urine formation in the animals possessing metanephridia. Now the kidney itself originates from the remnants of the coelom, which is very much reduced in modern mollusks. It is to be expected that filtration would be retained as a first step in urine formation, and this is found in all the mollusks so far studied. In the bivalve, or pelecypod, mollusks, the filtration is actually through the walls of the heart; in the gastropods the filtration is probably from capillaries in the walls of the kidneys; and in cephalopods the filtrate appears to be formed by special organs, the branchial heart appendages. Filtration pressures are not high, but the mollusks generally have a low protein content in their blood, and so the reabsorptive forces which oppose filtration are not high either. As a result, surprisingly large volumes of filtrate have been obtained in the few studies made on the subject. From the pericardial space, or branchial heart appendage, the filtrate passes next through the renopericardial canal into the kidney proper where two processes are simultaneously at work. Because the process of filtration cannot be finely discriminating, it merely separates water and solutes from blood proteins and cells. Many useful solutes remain in the filtrate and would be lost from the body, but they are absorbed by the cells of the kidney from the urine as it passes over the many folds into which the lining of the kidney is thrown. At the same time other cells of the kidney actively secrete materials into the urine. These three processes, *filtration, reabsorption, and secretion*, occur repeatedly in phylum after phylum regardless of the origin of the kidney tissues.

Mollusks may be exposed to desiccation during tidal excursions or at certain seasons of the year. A low metabolic rate and the usual presence of an impermeable shell make them well suited for surviving prolonged drought. The kidneys may release no urine whatever during such a period, the chief nitrogenous excretion, uric acid, simply forming crystals within the kidney or in the kidney cells. At the next wet period normal function is restored.

Arthropoda. Of the very large phylum Arthropoda only two major subgroups are treated here, the crustaceans and the insects.

Crustacea. The crustaceans which have been most studied originate their urine at a thin-walled sac held open by strands of connective tissue passing from the external surface of the sac to adjacent parts of the exoskeleton. Pressure in the blood which bathes the external surface of the sac or in a network of fine capillaries in its walls results in formation of a filtrate upon which other changes may later be made. The filtrate formed in the end sac runs next through a long canal lined with cuboidal or columnar epithelium before it is voided to the exterior. While the filtrate passes through the canal, salts and water may be reabsorbed to conserve them for the body, while other materials are secreted into the urine and thus excreted at a higher rate than they would have been by the process of filtration alone. See CRUSTACEA.

Insecta. Excretion in the insects follows an entirely different pattern from any described to this time. In his early work on the histology of the silkworm, M. Malpighi described the tubes which were later named the Malpighian tubes, when their excretory function was learned. These tubes, 2–150 in number, lie in the abdominal cavity; when they are few in number they are long, and when

are more numerous they are short. The Malpighian tubes begin development as blind pouches, lie free in the body cavity, convoluted or not depending upon their length, and finally open most frequently into the midgut but sometimes into the hindgut. There is no obvious mechanism for filtration of blood, and it is thought that the urine in all the members of this large group of animals is formed entirely by secretion of the cells lining the tubes. In fact, the tubes sometimes secrete other than excretory substances; for example, silk for the cocoon is secreted by the Malpighian tubes of some of the Collembola and Neuroptera. In the absence of filtration it might be surmised that there would be no need for reabsorption, but it has been shown that uric acid is secreted in the form of the relatively soluble sodium and potassium salts in the upper segment, and that the free uric acid is precipitated in the lower segment where the water and the liberated bases are reabsorbed and used again. In insects, much reabsorption of water occurs in the hindgut. *See* INSECT PHYSIOLOGY; INSECTA.

Echinodermata. The phylum Echinodermata has been restricted to the sea, perhaps in part because its members did not develop an excretory system of competence and flexibility. The departure of the body fluids from the composition of the surrounding sea water is small and in detail controversial. The animals, therefore, face few problems with respect to the regulation of salts and water. Waste materials presumably diffuse quickly from the body spaces into the surrounding sea water. When experimental injections of particulate matter or of dyes are made, the material is usually picked up by wandering ameboid cells, the coelomocytes, and delivered into the digestive tract from which they are ultimately extruded with the feces. The water-vascular system of the animal is not a rapidly circulating stream adapted to meet the needs of a circulatory system but is principally the means of renewing hydraulic fluid which is used in the operation of the many tube-feet of the animal. Thus in this large and ancient group of animals there is an absence of the standard processes of filtration, reabsorption, and secretion which are characteristic of most of the other invertebrate animals. But at the same time concomitant limitations in body size, degree of activity, and freedom from the environment may be noted. *See* ECHINODERMATA.

VERTEBRATA

On embryological grounds the Echinodermata are thought to be one of the groups most closely related to the vertebrates, but while the echinoderms have no excretory structures, the kidney of the vertebrate is a well-formed and effective organ. To add complexity to the view of relationships, the cephalochordates, thought to be close to the primitive line of the vertebrates, possess a nephridium, the primitive kidney of the lower invertebrates.

The nephron. With minor exceptions, each class of the vertebrates possesses a kidney made up of units like those of its relatives. The unit structure is the nephron, and these nephrons differ in detail

evolved animal by many nephrons, so that a single kidney of the human may contain 1,250,000 of these units. *See* KIDNEY.

A nephron may be said to consist first of a Malpighian body consisting of a tuft of capillaries, and the glomerulus, surrounded by a close-fitting sac, Bowman's capsule. Blood pressure in the tuft of capillaries filters a large amount of fluid into the capsule from whence it passes into a much-convoluted tubule draining the fluid away through a narrow passage between closely packed kidney cells. These cells carry out the complicated processes of reabsorbing for the body the materials of usefulness and of adding to the urine, from the blood, materials of such a nature that special mechanisms could usefully be developed for their transportation across the kidney cells. Nevertheless the utilization of the three processes of filtration, reabsorption, and secretion used throughout the rest of the animal kingdom can be noted. A few special features of excretion in the different vertebrate classes are worthy of comment.

Fish. In the Insecta there is a specialization of structure such that secretion is the primary mechanism used by the Malpighian tube in fulfilling the function of excretion for the animal. A similar device has been developed by some marine fishes; the Malpighian body of the vertebrate nephron is gone, leaving only the tubule, closed at the end where the glomerulus should be found. Since the means of accomplishing filtration is gone, it is clear that secretion is the only remaining mechanism for accomplishing the excretory functions of the kidneys in these animals. These nephrons make up what is called an aglomerular kidney, and the aglomerular fishes represent a successful group of animals, being numerous and widely distributed. It is thought that the reduction or loss of the glomerulus represents a means of water conservation. Fish blood is osmotically inferior to the salt-water environment. The fishes consequently continuously lose water osmotically through the gills and pharyngeal membranes, and any means of reducing excretory water loss has survival value.

The general osmotic problem faced by marine fishes is of such overriding nature that an entirely extrarenal excretory mechanism has developed to meet it in the gills of fishes. It may be recalled that the characteristic cation of living cells is potassium, whereas the abundant cation of body fluids and saline waters is sodium. Cell membranes are permeable to these small ions so that a majority of all animal cells continuously pump sodium out of the cell and potassium into the cell and must do so to continue normal life. Certain cells of the fish gill direct the pumping action not generally outward but polarize it so that sodium entering from the

blood and tissue fluid is pumped against a steep gradient into the sea water. In this way the kidney of the fish is spared the task of excreting the sodium ion, a task which the kidney of most terrestrial animals must perform. *See* RESPIRATORY SYSTEM.

Amphibia. Amphibians are characteristically bound to bodies of fresh water. Their osmotic problems are therefore quite different from those of the marine fishes just discussed. Their nephrons begin the process of urine formation with vigorous filtration. As the filtrate passes down the tubules, more salt than water is reabsorbed, water ordinarily being present in excess, and a hypotonic urine results. Even so, salt is continuously being lost in the urine and so is at a premium in these animals. The deficit is made up by the skin, which has the power of taking sodium and chloride ions from the environment into the blood against a steep gradient.

Reptiles. The discussion of excretory specialization in reptiles will be deferred until the section dealing with the excretion of nitrogenous wastes.

Birds. Marine birds face a situation similar to that of the marine fishes. Although they may not lose water by osmosis directly to the sea water, some of them remain in environments where fresh water is not available. At the same time they may feed on marine invertebrates whose body cells are in osmotic equilibrium with the sea water, and so through the digestive tract they are exposed to salt loads that constitute a serious problem to the body. In these cases recent research shows that it is not the kidney which has assumed the excretion of salt at a higher rate, but special glands of the head region, the nasal glands. These glands are able to produce a salt solution very hypertonic to the blood, which is blown, drips, or is shaken from the beak. They are thus able to meet the imposed load easily. The function of these glands has been well integrated into the physiology of the animal and is controlled from the osmoreceptors of the body via the central nervous system. *See* SALT GLAND.

Mammals. Desert mammals meet somewhat the same excretory problem that marine fishes and some marine birds must meet, namely a limitation on the water available for excretion. It has been shown that their kidneys have developed the capacity to excrete salt far beyond that of the ordinary mammal. These animals can comfortably eat dry food and drink sea water and, because they excrete a urine hypertonic to sea water, maintain themselves in satisfactory water balance. Some marine mammals are not only limited in their fresh-water intake but eat marine invertebrates containing large amounts of salt. These animals show the same adaptation as the desert mammals: great ability to concentrate salt in the kidney tubules.

Excretion of nitrogenous wastes. Nitrogenous wastes, the group of compounds which, it may be assumed, necessitated the development of excretory systems, are ammonia, urea, and the purine derivatives. From the standpoint of the enzymologist

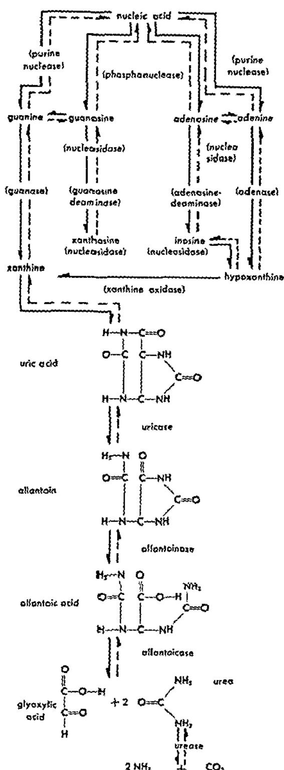


Fig. 1. The sequence of breakdown from nucleic acid to ammonia. The reverse pathways are dotted to indicate either a lack of knowledge of the mechanism or that other enzymes than those shown are involved. (Modified from M. Florkin, Biochemical Evolution, Academic Press, 1949)

these compounds make up a single family, all the members of which can be synthesized by well-known enzymes starting either from the ammonia generated from amino acids or starting with purines generated from nucleic acids. The intermediates are illustrated in Fig. 1. It has been estimated that 95% of the nitrogenous wastes of the average animal comes from proteins and only 5% from nucleic acids, but because of the metabolic interconnection one cannot predict from these percentages what the form of excretion will be. Instead it turns out that the habitat and the need for conservation of water determine the form of excretion. In the animals which spend their lives in an abundance of water—for example, amphibian larvae and many marine invertebrates—the ammonia split from protein is simply liberated to the water and carried away. The animals are called ammonotelic. But these animals also produce nitrogenous wastes from nucleic acids. Instead of liberating adenine and guanine directly to the environment, the complement of enzymes shown in Fig. 1 is invoked, the material is split all the way to ammonia, and the ammonia released therefore originates from protein and nucleic acid. This general principle of a common denominator of nitrogen excretion was enunciated by M. Florkin. The results of the principle, and some exceptions, may be seen in Fig. 2.

Elaboration of urea. Animals of larger size, in particular those which might not have a steady supply of water, cannot tolerate the accumulation of ammonia which would result from the use of ammonia as the single nitrogenous excretory product, and it was necessary that a less toxic material be substituted. In retrospect it is clear that the next evolutionary step, the elaboration of urea, was both most important and very difficult. Early biochemists speculated that urea was formed from two molecules of ammonium carbonate in what appeared to be a relatively simple synthesis. Progress in working out the actual pathways was slow in spite of the fact that the action of the enzyme arginase in splitting urea from the amino acid arginine had long been known. There was simply not enough arginine available to account for the amounts of urea formed. It was through the genius of H. A. Krebs that the solution was found. Only catalytic amounts of arginine need to be present. Arginase splits off urea, leaving a molecule of ornithine. This is condensed with an ammonia and a carbon dioxide molecule to form citrulline, which in turn is condensed with another ammonia to form arginine again. Ornithine and citrulline had long been known in nature as constituents of a few plants, but their possible roles in such a cycle had not been suggested until Krebs demonstrated that catalytic amounts of these compounds greatly accelerated the in vitro synthesis of urea by liver tissue. Since that time, much detailed information has been accumulated about this reaction, including the energy sources needed for the synthesis. Unlike the simple liberation of ammonia, the synthesis of urea demands the provision of a special complement of enzymes and the utilization of some of the body's supply of energy. The result for the whole animal, however, is the substitution of urea, a relatively soluble and nontoxic material, for the highly toxic ammonia molecule. Animals which synthesize and excrete urea are spoken of as being ureotelic. Most fishes and many amphibians fall into this category. There is a decided tendency for the purine bodies to be degraded to the level of urea and excreted in this form (Fig. 2). See NITROGEN EXCRETION.

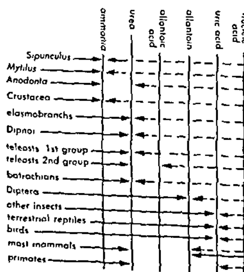


Fig. 2. Illustration of the principle of excretion of a common nitrogenous waste whether derived from protein or nucleic acid. (Modified from M. Florkin and G. Duchateau, Arch. Intern. Physiol., 53:267–307, 1943)

Urea synthesis and excretion. The synthesis and excretion of urea made animals independent of a direct and abundant source of water. The solubility of urea, which appears actually to have been one of its virtues for the development of the primitive kidney, turns out to be a disadvantage to animals like the birds and desert lizards which, for conservation of weight and water, need an excretory material which is both nontoxic and insoluble. At the same time the material must be capable of being secreted by kidney cells in solution in water and precipitated only in regions from which it can be expelled safely from the body. From the evolutionary point of view such a material had been available all the time to the more primitive animals. It will be recalled that in the fish and amphibians (compare Fig. 1) the nitrogen from the purines was degraded via uric acid to urea. The birds and most reptiles arrest this degradation at the uric acid stage. By enzymatic processes not yet fully understood, they carry the synthesis of the protein ammonia past the stage of urea synthesis and all the way up to uric acid. The uric acid is precipitated in the ureters and cloaca, and the pasty material is eliminated via the rectum. In this way another great step is made in the conservation of water. These animals are called uricotelic. It may

be noted from Fig. 2 that such a mechanism was independently developed by the insects, and for the same reasons. It should be pointed out that when such mechanisms develop there is not necessarily a clear-cut departure from the previous condition. Careful study shows therefore that various animals excrete varying mixtures of ammonia, urea, and uric acid, instead of only a single material.

Remaining problems. Many details are known about the excretory processes in man and lower animals. One of the great remaining problems of biological science is elucidation of the family of mechanisms by means of which secretion takes place. Cells are obviously able to move water, salts, and organic compounds of many sorts against steep concentration gradients. Although the energy expended can be calculated and the source of metabolic energy can be poisoned, none of the chemical or physical steps involved in the transport is known. [A.W.M.]

Bibliography: M. Florkin, *Biochemical Evolution*, 1949; M. Florkin and G. Duchateau, *Arch. Intern. Physiol.*, 53:267–307, 1943; E. S. Goodrich, The study of nephridia and genital ducts since 1895, *Quart. J. Microscop. Sci.*, 86:115–392, 1945; J. R. Robinson, *Symposium Soc. Exp. Biol.* 8:42–62, 1954

Exfoliation, rock

The splitting off of concentric thin sheets or shells from the surface of massive rocks is a common weathering process in moist climates. Both above and beneath the ground surface, hydration of complex silicate minerals through the action of percolating water carrying minute amounts of carbon dioxide causes peripheral expansion in rocks. The process is particularly effective at edges and corners of joint blocks where penetration can be from several directions. Separation of successive curved shells tends to produce increasingly rounded forms. Chemical exfoliation is distinct from the spalling of rocks by fire or sheeting of rocks in response to release of load. *See* WEATHERING PROCESSES.

[C.F.S.S.]



Exfoliation. Lower Quarter Dome, Yosemite Valley, California. (USGS)

Exopterygota

A division of the subclass Pterygota of the class Insecta. This group of insects is also referred to as the Hemimetabola. The type of metamorphosis provides the basis for this name. These insects have an incomplete, or direct, metamorphosis in which the young, called nymphs, usually resemble the adults in structure. Those immature stages which are aquatic are termed naiads, while the imagens, or adults, are aerial. *See* INSECTA; PTERYGOTA. [E.O.E.]

Bibliography: A. D. Imms, *A General Textbook of Entomology*, 9th ed., 1957.

Experiment

The test of an hypothesis under controlled conditions. The experiment is one of the distinctive tools of the scientist. It enables him to put his questions to nature and receive answers. These answers lead him to new problems whose solutions require more complex experiments (seeking smaller differences), improved techniques, detailed plans, and better analysis of results.

Results of experiments are usually recorded numerically. Statistical methods are used to reduce the data to summary forms with estimates of the magnitude of such effects as averages, variances (dispersal of the data around the average), and the relationships between the variables being measured. Associated with these estimates are tests of significance that permit the researcher to go beyond the samples and make inferences about the characteristics (parameters) of the populations from which the samples are drawn. Statistics helps the researcher understand the magnitude of the imperfections of his experimental data. *See* BIOMETRICS; STATISTICS.

Experimental error. Results of experiments are affected not only by the "treatments" (experimental procedures whose effects are being measured and compared) but also by certain extraneous variations that tend to mask the effect of these treatments. Two main sources of experimental error which must be distinguished are (1) an inherent quality of the experimental material (the material on which the treatments are acting) and (2) lack of uniformity in the conduct of the experiment or failure to standardize techniques.

The presence and cause of experimental error need not concern the investigator, provided his results are sufficiently accurate to permit definite conclusions. Often, however, the results of

ments are greatly influenced by experimental errors, making decisions difficult. *See* ANALYTICAL CHEMISTRY, ERRORS IN.

Planning experiments. Inferences that can be made from results depend upon the way the experiment is conducted; thus planning of an experiment must include a detailed description of the proposal, together with such considerations as the following.

Objectives of the experiment. The purpose of the experiment must be clearly defined in terms of questions to be answered, hypotheses to be tested, specifications to be met, or effects to be estimated. The statement should indicate the extent to which inferences from the data will be applied.

Usually a sample is used. The size (that is, the number of units analyzed) of the sample needed depends upon the accuracy desired for the results, the variability of the experimental material, the errors of measurements, the magnitude of the differences to be measured, and the time and money available for the study. The results of any piece of research are no better than the sample used. A sample is interesting only if it furnishes information about the population to which the conclusions are to be applied

Description of the experiment. After the objectives have been determined, the researcher should describe the following aspects of the experiment.

1. Treatments. The treatments whose effects are to be tested and measured are selected. It is important to understand how these treatments will help reach the objectives of the experiment. The choice of treatments may also have a substantial effect upon the precision of the experiment. Often it is desired to test many varieties, levels of chemicals, sources of vitamins, or kinds of spray material. These treatment combinations make up single-factor experiments. Striking gains in precision may be achieved in factorial experiments, where several factors are investigated simultaneously. Such experiments are useful in exploratory work to determine whether factors have any effect, to find interactions among factors, and to permit recommendations that apply to a wide variety of conditions.

2. Experimental material. The experimental material is chosen and its inherent variability is evaluated. [illegible] mir[illegible] var[illegible] the [illegible] extreme, it may be unwise to select uniform material because the responses obtained may not apply to the regular, unselected material.

The term experimental unit denotes the group of material to which a treatment is applied in a single trial of the experiment. The unit may be a patient in the hospital, an electronic tube, a group of pigs, a tree, or a missile. It is the characteristic of such units that they produce different results even when subjected to the same treatment.

3. Size of the experiment. This factor helps determine the sensitivity of the experiment, that is, the size of the detectable differences between the effects of treatments. One method of improving sensitivity is to increase the size of the experiment. This can be done by using more replications, that is, by increasing the number of times a complete set of treatments is run, or by using larger experimental units. Whatever the source of variation, replication of the experiment decreases the error associated with the difference between the average effects of two treatments, provided the precaution of randomization within replications has been used and that inclusion of additional experimental units or larger ones has not introduced more variation. Statistics and the scientist's experience can assist in determining the efficient size for the experiment.

4. Experimental techniques. The principal objectives of good techniques are to secure uniformity in applying treatments, to impose sufficient control over external influences so that every treatment produces its effect under comparable and desirable conditions, to provide unbiased measures of the effects of treatments, and to prevent gross errors. Refining the techniques helps to reduce the experimental variation

5. Related variables. These variables which predict to some extent the performance of the experimental units if measured and used will increase the efficiency of the experiments. Analysis of covariance enables the experimenter to estimate from the data the extent to which the results were influenced by variation in these supplementary variables.

Experimental designs. Experimental error often can be minimized by choosing an efficient design, a set of rules for allocating treatments to experimental units. Each restriction imposed by the experimental design has a definite purpose. Of the many types of designs available only the basic ones and a few of the variations are to be mentioned.

Completely randomized design. In this type, the treatments are allocated to experimental units entirely by chance. These units should be handled in random order at all stages of the experiment at which order is likely to affect the results. This design is desirable for laboratory research, especially in physics, chemistry, bacteriology, or experimental cookery, in which mixing produces quantities of homogeneous material that can be tested under uniform conditions.

Randomized complete block design. In this design the experimental units are in compact or homogeneous groups, each group containing enough units for all treatments. Usually several groups (replications) are needed to give an estimate of the effects being tested. During experimentation a uniform technique should be employed for all units in a group. The experimental error arises from variation within groups.

Latin square design. In this design, the treatments are grouped into replications in two ways, thus providing an opportunity for two reductions in error by skillful planning. The location of leaves on a plant and the difference between plants may determine the basis for grouping.

Incomplete block design. This design is used for experiments in which a large number of treatments is to be investigated and the number of homogeneous experimental units which can be grouped is small. These designs are composed of blocks or groups of experimental units which are smaller than a complete replication.

An example of how designs are selected to fit into experimental situations is given. In tests of mosquito repellents which involve exposure of treated arms to mosquitoes, the block consists of two arms of a subject at one time. To test six repellents the design in the table provides for each of five subjects (A, B, C, D, and E) to submit his two arms to treatments three times.

Incomplete block design

Day	Individuals									
	A		B		C		D		E	
1	1	2	1	3	6	2	3	6	4	5
2	6	5	4	6	4	1	5	1	2	3
3	3	4	5	2	3	5	2	4	6	1

Every possible pair of treatments occurs once on some individual on one day. Repellents 1 and 2 [illegible]

mitted his arms to all six repellents.

Incomplete block plans such as the one illustrated often are 20–100% more sensitive than randomized complete block designs. That is, smaller differences can be determined in an equivalent number of trials.

Split-plot design. This design is used for experiments in which an extra factor is introduced by dividing each experimental unit into two or more parts. Any of the plans presented above could have each unit divided into parts permitting the use of an extra factor. In industrial experimentation, frequently one series of treatments requires rather large bulks of experimental material, such as types of furnaces for the preparation of alloys. Another series, such as the molds into which the alloy is poured, can be compared through use of smaller units.

Factorial experiment. In factorial experiments more than one factor is included in the treatment combinations. The quantitative or qualitative factors may be either independent or not independent.

If factors are independent, these factorial experiments may be arranged in completely randomized, complete, or incomplete block designs. As the number of factors or levels of each factor increase, the number of treatment combinations increases rapidly. When the number of homogeneous units in groups is limited, incomplete block designs are desired. For factorial experiments, a special series of incomplete block designs has been developed where information on certain treatments is sacrificed to secure better information on other treatments, or where higher-order interactions are completely or partially confounded with block.

If quantitative factors are used regardless of independence, one thinks of the yield or response as a function of the level of the variables. Sequential experimentation is used to estimate the optimal point on a response surface and to explore the nature of this surface near this optimum.

The limitations of experimental materials, laboratory facilities, or time and convenience of investigators often dictate the specifications of the design, which is a logical plan. An adequate experimental design involves not only a satisfactory plan for conducting the experiment, but includes appropriate methods for evaluating results.

Analysis. The analysis of data calls for careful selection from the available efficient and flexible set of statistical methods. The use of these methods requires judgment and an understanding of the basic assumptions involved.

It should be emphasized that statistical analysis cannot increase the validity of data. The accuracy of mathematical inferences from data must be limited by the precision, accuracy, and adequacy of the measurements of the observations. If the experiment has been conducted so that treatments are confounded with extraneous effects, statistics cannot give reliable estimates of treatment effects. For well-planned and well-executed experiments, the sample of observations will provide an unbiased estimate of the effects, with measures of the uncertainty of these estimates.

Interpretation. The statistician should help the experimenter plan his experiment so that the results will answer definite questions. The experimental results should be presented in concise and organized form so that others can draw accurate conclusions. Drawing conclusions without the help of statistics is similar to building a suspension bridge without knowledge of the tensile strength of cables. A perfectly safe bridge can be built without such knowledge, but it will be unnecessarily expensive. [G.M.CO.]

Bibliography: W. G. Cochran and G. M. Cox, *Experimental Designs*, 2d ed., 1957; F. M. Wadley, Incomplete block design adapted to paired tests of mosquito repellents, *Biometrics Bull.*, 2(2):30–31, 1946.

Explosion and explosive

An explosion is a violent and noisy outburst, or sometimes a peaceful burning that would be better described as flame or combustion. This discussion, however, will be limited to those events associated with buildup and release of pressure. In this sense the term includes not only transient occurrences, such as the bursting of a shell or the burning of powder in a gun barrel, but also continuous, steady-state effects such as the expulsion of gases from a rocket nozzle. For discussions of special cases of explosion, *see* INTERNAL COMBUSTION ENGINE; MINING SAFETY; NUCLEAR EXPLOSION.

Explosion is usually associated with rapid chemical reaction. The reaction must evolve heat, and it must also produce some gaseous products. The explosive may be a single compound such as trinitrotoluene (TNT) or a mixture such as gunpowder (potassium nitrate, charcoal, and sulfur). All explosives give simple end products, such as CO_2, CO, H_2O, and N_2, these are the most stable atomic combinations at ambient temperatures. Oxygen-deficient explosives may also yield uncombined hydrogen and carbon. Certain transitory molecular fragments, or free radicals, such as OH, H, and O, are usually present in the high heat of the explosive reaction, but combine to form stable molecules when the gases cool.

A distinction can usually be made between the chemical events in the explosion and the external events that result from it. Different phenomena may cause similar, and even indistinguishable external effects. Thus the bursting of a large pressurized vessel, the explosion of a gas mixture in a building, or the detonation of a high-explosive bomb, may cause virtually identical effects at a distance. A remote observer could not know the true nature of the explosion. He might even refer to all three events as detonations; but, in the scientific sense, only the bomb produces a detonation.

Blast wave. When pressurized gas is suddenly released in a fluid medium (air or water), a pressure pulse travels outward. The speed is closely related to the sound velocity, about 1050 ft/sec in air and 5000 ft/sec in water. With high explosives, the pressure buildup is so fast that it creates a steep-fronted pulse called a shock wave, or, in air, a blast wave. The shock-wave velocity depends on the strength, that is, on the excess pressure at the wave front; it is always higher than the speed of sound, but approaches the acoustic level as the shock strength falls to zero (Fig. 1).

In air, the relationship between excess pressure at the wave front p_s and wave velocity D is given approximately by the formula

$$\frac{p_s}{p_0} = \frac{7}{6}\left(\frac{D}{c_0}\right)^2 - \frac{1}{6}$$

where p_0 is the atmospheric pressure, and c_0 is the sound speed. In water, because it is much less compressible, the shock pressure is 100–1000 times higher than the air blast pressure from a similar explosive charge at a comparable distance. The shock-wave velocity in water, however, deviates only slightly from sound speed.

Much of the explosion energy is concentrated in the blast wave. Part is in the form of compressional work, represented in mathematical terms by the quantity $\int p\, dv$. Some of the energy is heat, produced when the gas passes through the shock front; the rest is kinetic energy associated with the wind behind the shock. A large fraction of the blast-wave energy is transferred to successive layers of the air as the shock front advances, and, because the frontal area increases by the square of the distance, the energy is spread ever more thinly in space. For this reason, pressure, temperature, and wind velocity drop off sharply with distance from the explosion, and the damaging power diminishes accordingly.

The characteristics of shock waves are studied by electronic and optical devices. A pressure-sensing instrument for air blast first records the sharp pressure rise of the shock front, followed by a region of decaying pressure in the so-called positive phase of the blast. When air enters the shock front it is given a strong forward motion and is thus greatly densified. This compression largely accounts for pressure rise; therefore, the wind velocity and the shock pressure are closely related. The positive pressure region behind the shock front lasts for only a few thousandths of a second in the case of high explosive bombs, but remains for about one second in a nuclear blast. Behind it there follows a suction phase, in which the pressure dips below atmospheric level and the wind reverses direction. The seeming paradox of windows being blown outward by an explosive blast is due to this suction phase.

Underwater explosions create similar effects. However, the pressures are much higher, the water movement is slower, and the suction phase is unimportant. Additionally, there is a second compression wave of lower pressure and longer duration caused by a contraction and second expansion of the explosion gas bubble; in fact, several weaker bubble pulses are also observed.

Blast damage. When an air blast wave strikes a vertical wall, the air is brought to rest and piles

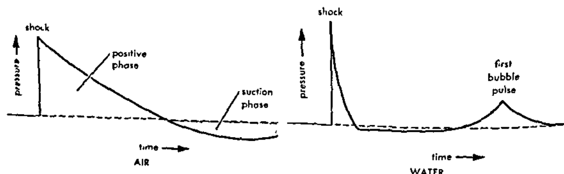


Fig. 1. Pressure versus time curve for shock waves in air and water.

up at the wall. Therefore, the instantaneous pressure on the wall is higher than the incident shock pressure. For acoustic waves of vanishing amplitude, the shock pressure is just doubled on impact, but for very strong shocks the excess pressure in the reflected wave may be as much as eight times the incident pressure.

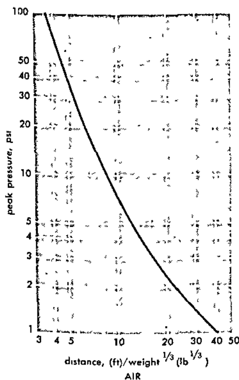
Under the force of an impacting blast wave, the wall begins to move. If it is massive and relatively flexible, that is, if the natural period of the wall is relatively long, the pressure may drop to zero before serious damage has been done. On the other hand, light and brittle members, such as window panes, may crack almost immediately. The destructive effect therefore depends on several factors, having to do both with the structure and with the blast wave.

For high-explosive bombs and industrial explosions, the damage-causing potential of the blast (*aside from window breakage*) *depends on the total* positive impulse, that is, on the area under the positive phase of the pressure-time curve. For nuclear explosions, the positive pressure phase lasts much longer than the natural period of most structural members. The damage caused by nuclear blast therefore depends only on the pressure level at the shock front. A similar criterion applies to window breakage from any but the smallest of explosive sources.

Similarity. The laws of fluid motion impose a similarity restraint on the manner in which the shock wave variables, such as pressure and impulse, change with distance from the explosion. The similarity principle states that when the size of the charge is increased, the scale of the shock-wave phenomenon increases in proportion. Thus, for a twofold increase in charge diameter, a given shock-wave pressure will be produced at just twice the distance from the charge. The time scale increases in a similar way. Because diameter is proportional to the cube root of charge weight w the scaling factor is just $w^{1/3}$. The scaling law is highly significant, for it means that if careful measurements are made for one size of charge, the blast effects from any other charge of that explosive can be predicted (Fig. 2).

Strictly speaking, the similarity law operates only for charges of the same explosive composition and of similar shape, but it can be applied with considerable accuracy even to widely different explosives without shape restrictions. The starting point for comparison is the energy equivalence, the common basis being the equivalent weight of TNT. For example, the bomb dropped on Hiroshima was ascribed an energy equivalence of 20,000 tons of TNT. With this single figure, the Los Alamos scientists made reliable estimates of nuclear blast characteristics from measurements for TNT charges 1,000,000 times less powerful. The scale factor in this case was just 100.

Window panes commonly break under a blast pressure of about 1 psi. According to the similarity



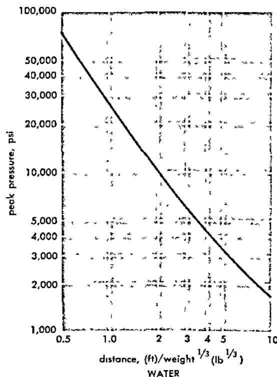


Fig. 2. Similarity curves for peak pressure as a function of charge weight and distance in air and water (based on TNT).

curve in the accompanying figure, such a pressure level occurs at about 40 ft from the explosion of 1 lb of TNT, and at about 14,000 ft, or 2.7 miles, from the explosion of a 20-kiloton nuclear bomb.

Energy release rate. Not all explosions produce a shock wave. For an unconfined gaseous explosion, the pressure rise is small and the energy is dissipated in a low-amplitude acoustic disturbance. It is a question of both reaction rate and scale; even a slow phenomenon on a sufficiently large scale can have all the destructiveness of a true detonation. For example, when the eruption of Krakatoa in 1883 spilled huge quantities of molten lava into the sea, the resulting steam caused a blast wave that shook the island of Batavia 96 miles away. A good measure of explosive violence might be the rate of total energy production divided by the surface area of the reacting medium.

Confinement can also be important. A burning gas mixture in the open air causes only a faintly audible whoosh, even when the flash of flame is quite spectacular. If the mixture is inside a large gasoline storage tank, however, the pressure builds up. Then, when the tank bursts and the hot gases erupt, a violent blast wave results. Here, the final rate of energy release to the atmosphere is high. *See* SHOCK WAVE.

Deflagration and detonation. The familiar burning mode is called deflagration or flame (*see* COMBUSTION; FLAME). Whether it produces an explosion depends on whether it is confined and causes the pressure to rise. Thus, when gasoline burns in an automobile cylinder it is said to explode, whereas in a blowtorch it is said simply to burn. There is, however, little basic difference between the two phenomena.

An important factor of deflagration is the burning rate. For the purpose of definition, the flame can be considered to be standing still, with the fuel mixture streaming into it and the burned products flowing away to the rear. The streaming velocity of the fuel mixture is called the burning velocity. It is characteristic, in the main, of mixture composition, but it also depends on temperature and pressure.

Burning velocities for gases ordinarily range from a few centimeters per second to about 1 meter per second. The flame advances by heat conduction and diffusion: fast-moving and chemically active molecules, often free radicals such as H and OH, diffuse from the burning zone into the unburned layer and start reaction. The layer-to-layer speed depends on the molecular velocities in the flame. Light, high-speed molecules, or molecular fragments such as hydrogen atoms, favor a high flame speed, as does a high reaction temperature. The chemical reaction rate also affects the temperature gradients and controls the heat flow.

The advance of a flame in an open-ended glass tube filled with a combustible mixture can be easily followed with the eye. Ordinarily it makes little noise. Under certain circumstances, however, a new phenomenon occurs. Depending on the kind of gas, the dimensions of the tube, and other factors, a quiet, slow-moving flame may suddenly develop into an ear-shattering explosion that pulverizes the glass in a most violent fashion. This transformation can take place even when the ends of the [illegible] allow free expansion for the burn- [illegible] ona- [illegible] ents [illegible] h a flame, the speed is incredible, one thousand or more times as fast. For example, an ethylene-air mixture burns with a velocity of 63 centimeters per second, and the detonation wave travels at a speed of 1734 meters per second.

The enormous difference in speed signifies a basic difference in mode of propagation. The flame advances essentially by layer-to-layer heat conduction; the detonation is carried forward on the crest of a shock wave. Deflagration is a diffusion phenomenon similar to the spreading of an odor in still air; detonation is a wave phenomenon similar to the propagation of sound.

Detonation develops from flame by a complex series of events that are not always precisely repeated in similar experiments. The transition is associated with tubes and is not known to occur in free space. In a tube, the expansion of the burned gas behind the flame can take place only by pushing away the gas in front (and also behind, if the tube is open at the rear; build-up to detonation is faster, however, in a tube closed at the rear). The gas flow caused by the expansion tends to become turbulent, and this speeds the burning rate. Small shock waves shoot out ahead of the flame, possibly because pockets of unburned gas are trapped in the turbulent wake. The final stage of transition to detonation is astonishingly sudden. The details of the process are complicated and little-understood. In contrast, the final, fully developed detonation has an elegant simplicity that sets it apart from other combustion effects (Fig. 3).

The impressive feature of detonation is its constant velocity. Once established, the speed does not change by more than 1 or 2 parts per thousand unless the mixture composition is nonuniform or the tube has irregularities. Changes in initial pressure and temperature have little effect. The detonation velocity is a characteristic constant for a combustible gas not unlike the velocity of sound for a noncombustible gas.

Detonation theory. To derive the detonation conditions for an explosive mixture, the wave front can be considered to be standing still. Consider two planes, one in the unburned gas in front of the wave and the other in the burned gas at the rear (Fig. 4). Matter under conditions p_1, v_1, E_1, and velocity u_1, is streaming toward plane 1; and matter under conditions p_2, v_2, E_2, and velocity u_2, is streaming away from plane 2. The symbols, p, v, and E represent pressure, specific volume, and specific energy, respectively. Three equations can express the following facts: (1) the mass flow rates,

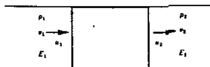


Fig 3. Microsecond flash photographs of a detonating stick of solid explosive. (a) Radiograph (x-ray). (b) Photograph. The radiograph shows the densification of the material immediately behind the detonation front and the rarefaction that sets in later as the gases expand. The outer cone of light in the photograph is the aerial shock wave, which is luminous at these very high pressures. (*From M. A. Cook, The Science of High Explosives, Am. Chem. Soc Monograph 139, Reinhold, 1958*)

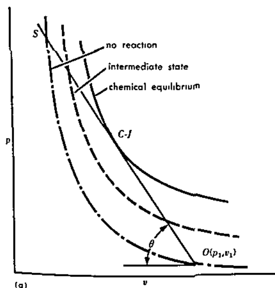
Fig. 4. Physical model of detonation front.

into 1 and out of 2, are equal (conservation of mass); (2) the difference between the total energy, internal and kinetic, flowing in at 1, and that flowing out at 2, is equal to the net work done by the gas (conservation of energy); and (3) the rate of momentum change between planes 1 and 2 is equal to the pressure difference (conservation of momentum). In addition, there is an equation of state that describes the particular properties of the detonating substance; this is a functional relationship, often not expressible as an explicit equation, between the final values p_2, v_2, E_2, and the initial values p_1, v_1, E_1. This equation involves the gas law and also reflects some assumption about the chemical state of the burned gas at 2. In all, since the initial gas conditions p_1, v_1, and E_1 are known, there are four equations with five unknowns: u_1, p_2, v_2, E_2, and u_2. Hence there is one degree of freedom, and an arbitrary choice of one of the unknowns will fix all five.

The solution to these four equations is usually represented by a functional relationship between p_2 and v_2, the pressure-volume curve for the burned gas. This is called the Rankine-Hugoniot curve or, simply, the Hugoniot.

Detonation wave velocity. Actually, the derivation outlined above is quite general. It applies equally to gases, liquids, and solids, not only to those that react chemically, but also to those that do not so react. The Hugoniot is, in fact, the general equation for a shock wave. Its form depends on the equation of state and on an implicit assumption about the chemical state of the medium that emerges in plane 2 If there is no chemical reaction, the Hugoniot passes through the initial point p_1, v_1; if there is, the curve lies above it.

The final detonation state lies on the "equilibrium" Hugoniot in Fig. 5*a*. Any point on this curve could represent the final detonation conditions and meet the four physical requirements invoked to derive the Hugoniot. Thus a whole spectrum of detonation states seems to be possible, but only one state is actually observed. To fix the detonation state, a fifth condition must be found. This is the so-called Chapman-Jouguet hypothesis. The detonation state, according to this principle, is uniquely represented by the point *C-J* on the Hugoniot in Fig. 5*a* This is the point of tangency of the chord



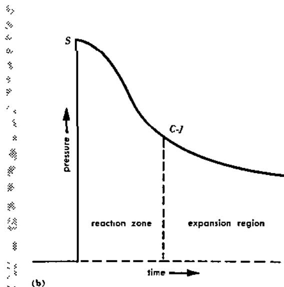


Fig. 5 (a) Hugoniot curves for a detonating gas. (b) Theoretical pressure versus time curve of detonation wave.

Calculated and observed detonation velocities

Composition	Calculated velocity, m/sec	Observed velocity, m/sec
Gases (by moles)		
$2H_2 + O_2$	2806	2819
$2H_2 + O_2 + 5N_2$	1820	1822–1810
$C_2H_2 + O_2$	2960	2920–2961
$C_2H_2 + O_2 + 4N_2$	2020	2015
C_2H_2	2070	2135–2160
Liquids		
Nitroglycerine	8060	8000
Nitroglycol	7630	8000
Solids (density in g/cm^3)		
TNT (1.50)	6480	6700
TNT (1.00)	5060	4900
TNT (0.50)	3730	3200
Tetryl (1.50)	7550	7300
PETN (1.50)	8150	7600

drawn from $O(p_1,v_1)$. Although the choice of the C-J point was originally an ad hoc hypothesis, the C-J state can be shown to be the only one on the Hugoniot for which the wave is steady.

According to the derivation, the detonation velocity D is equal to $v_1\sqrt{\tan\theta}$. The C-J velocity can be shown to be equal to the mass flow rate plus the acoustic velocity in the burned gas. The stable detonation therefore has the same speed relative to fixed coordinates as a sound wave in the burned gas. Actually, transitory detonation states are possible and are observed, that correspond to other points on the Hugoniot, but the C-J detonation is the only one that is stable.

Since, in computing the Hugoniot, any given chemical state can be postulated at plane 2, it is possible to erect a series of Hugoniots for all chemical states from no reaction to final equilibrium. A further extension of the theory shows that intermediate conditions in the chemical reaction zone lie on the straight line between S and C-J in Fig. 5*a*. The point S is the condition at the shock front (no reaction).

Detonation mechanism. In this emerging picture of the detonation wave, it appears that there is a shock front S followed by a region of decreasing pressure in which the reaction occurs. The reacting system comes to chemical equilibrium at C-J. The pressure here is about one-half that at S. The pressure contour of a detonation wave is depicted in Fig. 5*b*. In the reaction region, from S to C-J, the contour has a permanent form, but in the expansion region, beyond C-J, the contour changes and the pressure gradients become less steep with time.

The true nature of detonation is now revealed. The wave is fronted by a shock that quickly heats the medium to high temperature. A rapid chemical reaction follows. The heat released in the reaction maintains the shock front at constant strength and velocity. Detonation can therefore be described as a type of self-propelled shock wave.

The difference between detonation and deflagration does not lie in the chemistry, because the reaction for a given gas mixture may follow essentially the same kinetic mechanism in both cases; the difference lies, instead, in the manner in which successive layers of material are brought up to the high reaction temperature. A deflagration will transform into a detonation if conditions are favorable for the buildup of a shock wave. A high degree of lateral confinement, as in tubes, and a rapid chemical reaction are the main factors favoring such a changeover, but the geometry of the container and the sound velocity in the medium are also important.

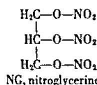
Detonation calculations. The Chapman-Jouguet equations provide a framework, but many theoretical building blocks are needed to complete the structure. For, in order to apply the equations, thermal data up to 6000°C are needed, and these are quite inaccessible by direct experiment. However, with the aid of statistical mechanics, such data can be computed from spectral measurements, that is, from the wavelengths of light that molecules emit and absorb. A further problem arises in the fact that detonation pressures for condensed explosives lie in the region of 200,000 atmospheres, far beyond the experimental range. To calculate the behavior of a gas at such pressures, it is necessary to rely on the theory of intermolecular forces. Indeed, the term gas as applied to the detonation state of a condensed explosive is only a formality, because the density is sometimes double that of liquid water. In spite of such difficulties, the theory has yielded reliable values of the parameters for all kinds of explosives. Where it is possible to check experimentally, the agreement is usually good.

The speed of the chemical reaction itself does not enter the Chapman-Jouguet theory. Indeed, it is possible to calculate detonation properties for materials that do not detonate at all. Presumably, the reaction is too slow. Other cases involve materials that detonate, but, because the reaction is sluggish, the velocity is below theoretical. Ammonium nitrate, for example, has a calculated velocity of 3460 m/sec, but, when it has been detonated (in a 9-in. diameter cartridge), the measured speed is only 1510 m/sec. Again, many explosives, such as nitroglycerine, exhibit a low-order detonation when poorly primed. The velocity in such cases may be only one-quarter to one-half of the C-J value. Thus an ideal detonation is in accord with the C-J theory, and a nonideal detonation, either because the reaction is slow or because it follows a different path, does not accord with theory.

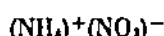
With a good rationale now available, development of new blasting compositions can be made on paper, with the elimination of much cut-and-try experimentation. A combination of theory and experiment is also yielding basic information about the behavior of molecules at very high pressures and temperatures.

High explosives. In an explosive mixture such as gunpowder, the fuel and oxidizer are in separate compounds; but in the pure high explosives, groups that serve these functions are fused together into

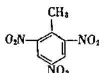
a single molecule. The oxidizer role in these explosives is played by the nitro group, $-NO_2$. In the aromatic nitro compounds such as TNT, it is joined to carbon; in the nitramines such as RDX, to nitrogen; and, in the nitrate esters such as nitroglycerine and PETN, to oxygen. Carbon and hydrogen supply the fuel for the reaction.



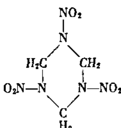
NG, nitroglycerine



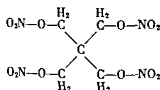
AN, ammonium nitrate



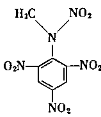
TNT, trinitrotoluene



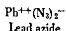
RDX, cyclotrimethylenetrinitramine (cyclonite)



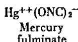
PETN, pentaerythritol tetranitrate



Tetryl, trinitrophenylmethylnitramine



Lead azide



Mercury fulminate

A fairly strict division can be made between explosives used for civil and military purposes. For many years following Alfred Nobel's epic discovery that nitroglycerine could be safely used and reliably detonated, nitroglycerine formed the base of the blasting dynamites. But its place has been increasingly taken over by ammonium nitrate, which is now the workhorse in a bewildering array of commercial explosives. Military explosives presently in use consist mostly of TNT, RDX, PETN, and tetryl.

Commercial explosives. Black powder was, from ancient times, and until Nobel, the only explosive It is not a high explosive, that is to say, a detonating explosive, for under most circumstances, it simply burns (or deflagrates). Black powder is made from charcoal, saltpeter (KNO_3), and sulfur. Potassium nitrate is the oxidizer, and the sulfur and charcoal make up the fuel. The manufacture of good powder requires skill, but it will not be described, for the importance of black powder as an explosive has almost vanished.

When Nobel discovered the use of nitroglycerine, the era of high explosives was opened. The shattering power of a detonating explosive when laid against rock or steel is in distinct contrast to the action of a deflagrating powder, which is completely ineffective unless confined. Nobel's discovery, therefore, created entirely new techniques of blasting and gave a great impetus to mining.

Ammonium nitrate, NH_4NO_3, has slowly risen to the place of major importance for blasting. In one sense, the trend would seem to be a reversion to the black-powder type of explosive, with NH_4NO_3 replacing potassium nitrate, KNO_3. Ammonium-nitrate explosives are similar to gunpowder. They contain the nitrate oxidizer and a fuel, which, in this case, may be almost any carbonaceous material from oat hulls to the sweepings of breakfast-food factories. For various reasons a bulky material is preferred To make a dynamite, either nitroglycerine or TNT is added to sensitize this mixture. The result is a cap-sensitive high explosive, quite different from gunpowder.

In coal mining, regulations require that an explosive must not be capable of setting fire to coal gas (methane). In effect, this restriction places a limit on the temperature of the explosive products. Explosives that meet this requirement are called permissibles.

In open-pit mining, an inexpensive explosive is now gaining wide use. This do-it-yourself material consists simply of prilled ammonium nitrate, an agricultural fertilizer, mixed with 5–6% furnace oil. The mixture is very insensitive, but in the wide and deep boreholes used for surface mining, it can be detonated with a booster charge of dynamite The earth-moving ability of this explosive is said to exceed that of dynamite.

Military explosives. TNT came into use as a military explosive during World War I, and still retains an important place. It is an oxygen-deficient explosive, and can therefore be improved and lowered in cost by the addition of ammonium nitrate, to make a mixture called amatol. Although this mixture was important in World War I, it is not used today. TNT derives its important virtues from its relative insensitivity to shock and its convenient melting point (81°C). Because it can be melted with steam, it can be safely cast into shells and other armament. In this respect TNT is unique, and, even though other explosives such as RDX are more powerful, TNT must still be used as the base for a castable mixture. The most common mixture is Composition B, which contains about 60% RDX, 40% TNT, and some wax as a desensitizer. Another mixture is Pentolite, a mixture of TNT and PETN.

RDX (cyclonite) rose to importance during World War II. In the early period of that conflict, a concentrated effort was made to develop a large-scale synthesis for RDX. This was successful, and the RDX plants soon began to turn out huge

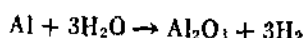
quantities. RDX is high-melting, fairly insensitive, and the most powerful compound in common use. Although slightly more energetic materials are known, RDX has almost the theoretical limit of stored energy in a molecule of this type. On decomposition, it is balanced to CO, as may be seen by writing the decomposition equation



$$C_3H_6N_6O_6 \rightarrow 3CO + 3H_2O + 3N_2$$

If there were sufficient oxygen to oxidize the CO to CO_2, a slight gain in energy might be possible. RDX is used in Composition B and, also, in conjunction with TNT, in the aluminized composition, Torpex.

By adding powdered aluminum to explosives such as RDX, a further increase in explosive energy can be produced. The aluminum reacts with the explosion products to produce aluminum oxide, for example



$$Al + 3H_2O \rightarrow Al_2O_3 + 3H_2$$

This reaction adds more heat to that already liberated by the decomposition of the explosive. There is, of course, an optimum amount of aluminum for any given explosive, at which the energy output reaches a maximum.

PETN is almost as powerful as RDX and is used interchangeably with it in some instances. The main use is in the detonating fuse, Primacord, a fabric-wrapped cord, which has a core of PETN and is widely used for commercial and military fuse trains.

Tetryl is a relatively sensitive explosive having excellent pressing properties. As a result it is common as the booster charge for military devices.

Detonators. The four explosives described above are called secondary explosives. They will ordinarily detonate only under the impetus of a powerful shock wave transmitted from a primary explosive. These latter materials, notably lead azide and mercury fulminate, are used in the form of detonator caps. When ignited by a fuse wire or by percussion, they go into detonation almost immediately. To further strengthen the shock wave from the detonator before it is transmitted to the main charge, a tetryl booster charge is normally employed. See DETONATOR; EXPLOSIVE FORMING; FUSE, EXPLOSIVE, PRIMER (EXPLOSIVE); SHAPED CHARGE.

Propellant burning in guns and rockets. A gun or rocket propellant does not detonate; or, if it does, the results are disastrous. The combustion mechanism in this case is deflagration. It is often called cigarette-type burning, to indicate that the propellant is steadily consumed by erosion at the surface. A flame stands at a short distance and heats the surface by radiation and conduction. The heating causes the surface material to evaporate and decompose. In this way, a stream of gas is produced to feed the flame and a steady state is set up.

The burning rate is reckoned as the rate at which the surface recedes. The rate depends on the gas pressure in the combustion chamber; in fact, for many propellants, the burning rate is just proportional to pressure. A representative speed is 2 cm/sec at 100 atmospheres.

The function of the propellant, in both guns and rockets, is simply to produce gas. The rate of gas production is roughly proportional to the product of the surface area of the unconsumed propellant and the pressure in the chamber. Therefore, since, in both rockets and guns, it is desirable to maintain a steady rate of gas production, the propellant grains are shaped so as to maintain a constant surface area as they burn. To produce this result in gun propellants, a perforated cylindrical grain is commonly used (see Fig. 6*a*). A grain of this shape, when it burns, both from the outside in and the inside out, tends to keep a constant area. In large, single rocket grains, on the other hand, burning is inhibited on the outer surface, and the grain burns only from the inside out. For this reason, a rocket grain usually has a star-shaped perforation in order to give it a large initial area.

The pressure in the gun breech rises rapidly when the powder is ignited. In a period of time t of about 1 or 2 milliseconds (1 millisecond = 1/1000 sec), it reaches a maximum value of about 50,000 psi. The outward movement of the bullet then causes the pressure p to drop. Its muzzle velocity depends on the integral $\int p\, dt$. An appreciable fraction of the energy of the propellant is used in accelerating the hot gases, as well as the bullet; ballistic designers therefore seek propellants with light (low-molecular-weight) gaseous products.

The propellant system of a rocket is designed to operate at constant chamber pressure and, therefore, constant thrust. It is very undesirable to have a propellant whose burning rate changes rapidly with pressure; if it does, instability results, and the pressure rises until the rocket explodes. Rockets therefore require special propellants with a low pressure coefficient in the burning rate equation. As with gun propellants, light combustion products are preferred. A common figure of merit for the efficiency of a rocket propellant is the specific impulse, I_{sp}; it is approximately proportional to

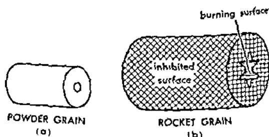


Fig. 6. Propellant grains. (a) A single powder grain which may have a diameter, or a web, as it is called, from less than 1 mm up to ¼ in. (b) Rocket grains which may be several feet in diameter, depending on the size of the rocket. Only a single grain is used in the rocket chamber.

$\sqrt{T/M}$, where T is the flame temperature and M the average molecular weight of the propellant gases. *See* PROPELLANT.

Besides black powder, which is mainly used in sporting rifles, the common gun propellants are either nitrocellulose or a mixture of nitrocellulose and nitroglycerine (double-base propellant).

Rocket propellants are of two main types: liquid and solid. The liquid systems use an oxidizer such as liquid oxygen or fuming nitric acid, and a fuel such as kerosine or hydrazine. Solid propellants may be of the nitrocellulose-nitroglycerine, double-base type (monopropellants), or mixtures (composite propellants) of an oxidizer, such as ammonium perchlorate, and a fuel, such as Thiokol.

Gas and dust explosions. Gas and dust explosions are serious industrial hazards, especially in coal mines (*see* MINING SAFETY). The severity of the effects depends on confinement, for gas explosions seldom result in detonation, and dust explosions almost never do. The burning is usually so slow that the pressure would not rise, except that the explosion gases cannot escape fast enough from the building or tunnel in which they are formed.

There are definite flammability limits for both combustible gases and dusts. Below a certain critical concentration of combustible, the air mixture becomes incapable of supporting a flame (*see* FLAME). In the case of gases, it is possible to provide warning of a rising concentration of combustible by means of a detector.

Coal mines, for example, make a regular practice of surveying the mine air for methane content. The minimum flammable concentration of methane is about 5%, but it can be readily detected at about 0.1%. The most common methane detector, which is also suitable for many other combustible gases, has a glowing platinum filament that catalyzes combustion of even slight traces of methane in the immediate vicinity of the wire. In a contaminated atmosphere, the wire becomes heated above its normal temperature, and the temperature rise can be detected by the change in the electrical resistance. Other methods of detection are also used.

Prevention. Unfortunately, in the case of dusts, there is no such convenient warning device Measures can be taken, however, to prevent dust explosions. Besides a simple improvement in housekeeping practices, these measures include the use of an inert dust, for example, stone dust, on ledges and other places where dust collects. If a sufficient quantity of the inert dust is present, it will stop the progress of flame in an otherwise combustible mixture. The action of the inert particles is to rob heat from the combustion and, in this way, to lower the temperature below the point at which the flame can spread.

Particle size is an extremely important factor in dust explosions. The minimum explosive concentration, that is, the minimum weight of dust in a cubic foot of air that will cause explosion, decreases sharply as the size of the particles is diminished. The total surface area of the dust particles is the underlying factor, and the area increases rapidly as particle diameter is reduced. More precisely, since the total weight is proportional to nd^3, where n is the number of particles and d is the average diameter, and since the total surface area is proportional to nd^2, the surface area for a given total weight of material will vary as $1/d$. As a result, a tenfold decrease in particle size leads to a tenfold increase in specific surface area.

Different gases vary in ease of ignition, and dusts vary in the ease with which the flame can be inhibited. For example, the most readily ignited hydrogen-air mixture requires only 1/20 as much energy for ignition as the most sensitive methane-air mixture. For dusts, it is easier to measure the amount of stone dust necessary to prevent inflammation. Enormous differences are found between dusts. Anthracite coal dust, for example, is barely flammable, and requires very little stone dust to prevent burning Cellulose-acetate resin, on the other hand, requires several times as much stone dust to stop a flame.

Dust explosions are very common. In some industries they occur almost routinely in the exhaust ducts, which are built to withstand the effects. The problem is aggravated by the fact that dusts generate static electricity, and, therefore, are sometimes self-igniting. Furthermore, when an explosion starts, it can spread by raising the dust in clouds ahead of the flame. Some metal dusts, notably magnesium and zirconium, are extremely hazardous.

Extinguishers. Gas flammability can also be reduced by the presence of certain inhibitors Most outstanding of these are the so-called alkyl halides (*see* FIRE EXTINGUISHER). The best known compound of this type is carbon tetrachloride, a common fire extinguisher, but the group of fluorinated compounds known as Freons is also very effective The action here is thought to be a specific chemical effect, possibly having to do with the halogen atoms that are broken away from the molecules by heat. [W.E.C.]

Bibliography: M. A. Cook, *The Science of High Explosives*, Am. Chem. Soc. Monograph 139, 1958; T. L. Davis, *The Chemistry of Powder and Explosives*, 1943; B. Lewis and G. von Elbe, *Combustion, Flames and Explosions of Gases*, 1951; J. Taylor, *The Detonation in Condensed Explosives*, 1952; U.S. Scientific Laboratory, *The Effects of Atomic Weapons*, 1950.

Explosive forming

The shaping or modifying of metals by means of explosions. The explosives may be of either the detonating or deflagrating type (*see* EXPLOSION AND EXPLOSIVE). Explosive gas mixtures or stored gas at high pressure may also provide the motive power.

Most types of explosive forming involve a mold into which a flat sheet of metal is pressed by the

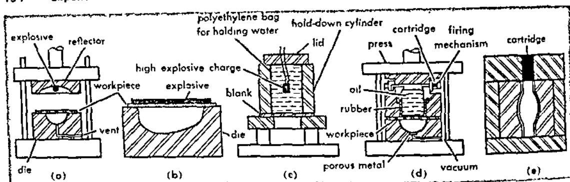


Fig. 1. Five methods of explosive forming (a) Shaped charge and parabolic reflector Shock and pressure are the motive power. Reflector and charge must be shaped to direct explosive force toward the workpiece (b) Flat high explosive is simply placed on top of sheet which lies on female die. Forming over a punch produces wrinkles. (c) The hold-down cylinder is often a carton of water or a bolted ring on lower die. (d) Gun-powder cartridges are usually applied in a press or enclosed die. This one uses oil and rubber to distribute force. Vacuum prevents air pocket which would retard action. (e) Bulge forming with cartridge power is comparatively simple and safe. Such devices properly vented are less noisy than many standard forming operations. (*From Explosives form space age shapes, in Steel, August 25, 1958*)

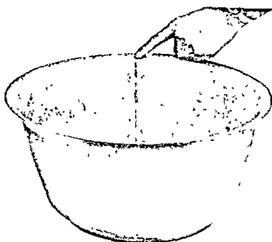


Fig 2. Typical metal shapes produced by explosive forming Note the weld in the bowl. (*National Northern Corp.*)

explosion. The metal is stretched uniformly by the explosive impulse. Even welds in the original blank survive the deformation without damage. The advantages of explosive forming over conventional forging methods accrue especially in the case of intricate shapes of which only a few items are required. Tooling costs are low.

Cold welds can be made between dissimilar metals by driving the two parts together under explosive impact. In other applications of explosive-forming methods, powders are pressed into solid billets.

In a quite different application, high explosives are used to cut large blocks of metal and even to split thin sheets into two layers of exactly one-half the original thickness. When an explosive charge is detonated in contact with the metal it produces a compression wave. On reflection from a free surface, the compression wave turns upside down to produce a tension wave. Along the planes in the metal where the tension in the wave front exceeds the strength of the metal, rupture occurs.

Explosives can also be employed to extrude metal shapes and to punch hard metals with the aid of dies. Shapes produced explosively are very exact and free from the fine cracks that sometimes result when pressure is slowly applied. Forces far exceeding those of the largest hydraulic presses can be applied by explosives.

Metals can also be hardened under explosive impact with results that compare favorably with those from slow cold-working methods. *See* METAL FORMING. [W.E.G]

Exponent

In mathematics, a symbol or number written to the right of and above another symbol or number to denote how many times the latter is to be multiplied by itself. For example, $7^3 = 7 \cdot 7 \cdot 7$, and $a^3 = a \cdot a \cdot a$. By use of this convenient mathemati-

cal device, the number 420,000,000,000,000 can be expressed unmistakably and more compactly as 4.2×10^{14}, which is read as, "four and two-tenths times ten multiplied by itself fourteen times" or, more properly, "four and two-tenths times ten to the fourteenth power." This abbreviated notation is particularly valuable in expressing the extremely large or extremely small numbers encountered in modern scientific work. For example, 0.000000143 may be expressed as

$$\frac{143}{1,000,000,000} \text{ or } \frac{143}{10^9} \text{ or } 1.43 \times 10^{-7}$$

Operations with exponents are governed by the following rules:

1. $x^a x^b = x^{a+b}$. Example, $x^2 x^3 = xx \cdot xxx = x^5$.
2. $(x^a)^b = x^{ab}$. Example, $(x^2)^3 = (x^2)(x^2)(x^2) = x^6$.
3. $(xy)^a = x^a y^a$. Example, $(xy)^3 = xyxyxy = xxxyyy = x^3y^3$.
4. If $y \neq 0$,

   $$\left(\frac{x}{y}\right)^a = \frac{x^a}{y^a} \text{ Example, } \left(\frac{x}{y}\right)^3 = \frac{x}{y} \cdot \frac{x}{y} \cdot \frac{x}{y} = \frac{x^3}{y^3}$$
5. If $x \neq 0$,

   $$\frac{x^a}{x^b} = x^{a-b}. \text{ Example, } \frac{x^5}{x^2} = \frac{xxxxx}{xx} = xxx = x^3$$
6. If $x \neq 0$ and $-n$ is a negative integer,

   $$x^{-n} = \frac{1}{x^n} \text{ Example, } \frac{x^2}{x^5} = \frac{1}{x^3} = x^{-3}.$$
7. If $x \neq 0$, $x^0 = 1$.
8. If p/q is any rational number expressed with positive denominator q, $x^{p/q} = (\sqrt[q]{x})^p$.

   Example, $x^{-2/3} = \sqrt[3]{x^{-2}} = (\sqrt[3]{x})^{-2} = \dfrac{1}{(\sqrt[3]{x})^2}$
9. If a is any irrational number, x^a is equal to the limit of the sequence, $x^{a_1}, x^{a_2}, x^{a_3}, \ldots x^{a_n}$, where $a_1, a_2, a_3, \ldots, a_n$, etc., are the one-place, two-place, ..., n-place approximations of a. Example, x^π is equal to the limit of the sequence, $x^3, x^{3.1}, x^{3.14}, x^{3.141}, x^{3.1416}, \ldots$

Any number which can be written in the form $a + ib$, where a and b are real numbers and $i = \sqrt{-1}$, is called a complex number. Exponents which are complex numbers are frequently encountered. For example, any number may be expressed as a power of $e = 2.718\ldots$, the base of the system of natural logarithms. Many quantities in natural phenomena are so expressed. Where e has a complex exponent, e^{a+ib}, by Rule 1, $e^{a+ib} = e^a \cdot e^{ib}$. The value of e^{ib} may be found by substituting ib for x in the expansion

$$e^x = 1 + \frac{x}{1} + \frac{x^2}{1 \cdot 2} + \cdots + \frac{x^n}{1 \cdot 2 \cdot 3 \cdots n} + \cdots$$

By separating the real and imaginary parts, the Euler equation is obtained, $e^{ib} = \cos b + i \sin b$. Thus, $e^{a+ib} = e^a(\cos b + i \sin b)$. See ALGEBRA.

[A.N.L.]

Exposure meter

An instrument used in photography to measure the brightness of the light from an object or scene. The earliest exposure meter was the nineteenth-century actinometer, in which the actinic (ultraviolet) rays in the light darkened sensitized bromide paper to give a measurable indication of brightness. Later developments were the visual exposure meter, in which the brightness of a scene was measured by comparison with a calibrated lamp having controlled variable light intensity; the optical type, which involved a calibrated neutral-density optical wedge that could be adjusted in the instrument until the light from the object shining through the wedge matched the surrounding light; and the extinction type, which involved a dial with a series of numbers of varying visibility, the last number to be visible representing a measure of the reflected light.

The exposure meter almost universally in use today is photoelectric in nature. This instrument, introduced in 1932, uses a photovoltaic cell usually of the selenium-junction type (see PHOTOVOLTAIC CELL). The cell aperture is baffled to narrow the field of vision to correspond to that of the lens of the camera with which it is being used. Selenium cells have a maximum sensitivity of about 100 microamperes per lumen at 5700 angstroms and adequately cover the visible spectrum (see ANGSTROM). The cell acts as a current generator whose output is indicated on a microammeter calibrated in one of the common exposure scales. From this reading the lens aperture and shutter speed can be adjusted to the light intensity and the type of photographic film being used. See PHOTOMETER.

[C.CO.]

Bibliography: C. B. Neblette, *Photography, Its Materials and Processes*, 5th ed., 1952.

Expressivity, gene

The degree of phenotypic expression of a gene in those individuals in which it manifests itself. For instance, the expressivity of the gene for brachyury ($T/+$) in the mouse can vary from a barely perceptible shortening of the tail to complete taillessness. In the case of bilateral characters, low expressivity often leads to asymmetrical manifestation. For example, in brachydactyly, a condition where one or more segments of the fingers are abnormally short or even missing, the numbers of fingers affected vary and the two hands may differ in this respect.

A high degree of expressivity of a gene is usually associated with high penetrance, and vice versa. Expressivity can be influenced by the environment, by accidents of development, and by the remainder of the genotype. See PENETRANCE, GENE. [H.GR.]

Extraction

A valuable method of separation which can be applied conveniently on either laboratory or process scale. Liquid-liquid extraction, the form most used in analysis, is a technique in which a solution is brought into contact with a second solvent, essentially immiscible with the first, in order to bring about a transfer of one or more solutes into the second solvent. The separations that can be achieved by this means are simple, convenient, and rapid to perform; they are applicable equally well to small and large amounts of materials, and they require only simple apparatus. The technique of extraction is broadly applicable to most inorganic and organic materials.

Applications. Extractions have been used in analytical procedures for most of the elements of the periodic table. A large number of extraction procedures have been incorporated into the colorimetric methods for trace amounts of the metals and for the isolation and radiochemical purification of radioactive isotopes. In addition, most large-scale separations of radioisotopes employ extraction.

In the fields of organic chemistry and biochemistry, where extraction has been used extensively for the isolation and purification of materials, the technique of countercurrent distribution has become an essential tool. Many commercial processes involving extraction are used for the fractionation of organic materials, one of the notable examples being the petroleum industry.

Distribution law. The behavior of solutes distributed between two immiscible solvents is best described in terms of the distribution ratio D, defined as the ratio of the formal, or the stoichiometric, concentrations of the solute in each of the two phases. Thus

$$D = \frac{|A|_2}{|A|_1}$$

where A_i refers to the formal concentration of the solute A in the appropriate phase as indicated by the subscript. When the solvent pair includes water and an organic liquid, D is defined as $|A|_o/|A|_w$.

In ideal cases, in which the solute is not involved in chemical interactions in either phase, the value of D is constant, in accord with the classical distribution law of Nernst, and is referred to as K_D, the distribution coefficient. In many cases, including the majority of inorganic (metal) extractions, chemical interactions do occur. These chemical reactions require the modification of the simple distribution law by the incorporation of the appropriate reaction equilibrium expressions into the distribution equation relating D to the experimental parameters.

The results of solvent extractions may also be described by $\%E$, the per cent extraction. This is related to D as

$$\%E = \frac{100D}{D + (V_w/V_o)}$$

where V_w and V_o refer to the volumes of solvents used.

Process of extraction. Solvent extraction processes may be considered to involve three stages: (1) chemical interactions involving the solute in the first solvent, usually water, (2) distribution of the extractable solute species, and (3) chemical interactions of this species in the second, or organic, solvent.

Most metal salts are strong electrolytes that are soluble in water and insoluble in organic solvents. Metal extraction requires the use of reagents whose major function is to transform the metal ion into an uncharged, extractable species. The formation of an extractable species can be accomplished through either coordination involving chemical bonding or ion association involving essentially electrostatic attraction. This aspect of the extraction process is of sufficient importance in metal extractions to warrant their classification on this basis.

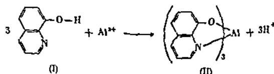
In the extraction of organic compounds, the nature of chemical interactions involving the solute is largely limited to proton-transfer reactions. Compounds such as phenols, carboxylic acids, and amines are generally extracted by organic solvents in the uncharged form, whereas as ions they remain in the aqueous phase.

The second stage of the extraction process lends itself to simple mathematical treatment inasmuch as the classical distribution law applies to each extractable species. However, elucidation of factors governing relative solubility in the two phases remains one of the most complex aspects of extraction processes.

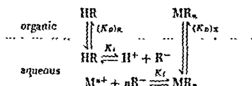
Chemical interactions of the solute or extracted species in the organic phase are important in ion-association systems because ion pairs have a tendency to form larger aggregates in the organic or relatively nonpolar phase. Polymerization may occur at higher concentrations. Organic phase reactions are unimportant in chelate extractions but are occasionally encountered with organic solutes such as the dimerization of carboxylic acids in non-hydrogen bonding solvents such as benzene.

Inorganic extractions. The major categories of metal extractions are chelate (coordination) and ion-association systems.

Chelate systems. Metal chelates are cyclic coordination compounds containing a metal atom in the ring. 8-Quinolinol (I), a chelating agent commonly used in extraction, contains two atoms (N and O) that will coordinate with metal ions to form a five-membered ring. Thus aluminum 8-quinolinate (II) is an uncharged chelate since its coordi-



nation number and charge have been satisfied by three 8-quinolinate anions. The uncharged chelate resembles many other unionized molecules by virtue of its extremely low water solubility and relatively high solubility in organic solvents. Many chelating agents, such as 8-quinolinol, may be represented as a weak acid, HR, in the following diagrammatic representation of chelate extraction.



It can be shown that D for such systems is

$$D = \frac{(K_D)_X K_i^n K_f}{(K_D^n)_R} \frac{(HR)_o^n}{(H^+)^n}$$

from which the extent of metal extraction is seen to be independent of the total metal concentration but to increase with the concentration of the chelating agent and the pH. Metal separations may be achieved by control of pH and oxidation state, and by the use of competing complexing agents (masking agents) which, by forming charged complexes, effectively alter the value of K_f and thus the extractability of the metal relative to others. Other typical chelating agents useful in extraction include acetylacetone, cupferron, dithizone (diphenylthiocarbazone), sodium diethyldithiocarbamate, and 1-nitroso-2-naphthol. *See* CHELATION.

Ion-association systems. Ion-association complexes involve the pairing of oppositely charged ions under the influence of electrostatic attraction. Ion-pair formation is favored in low-dielectric-constant media. Metals may be incorporated into either the cationic or anionic member of an extractable ion-pair complex.

Hydrated metal cations can couple with large organic anions such as those of the carboxylic acids or alkylphosphoric acids. Alternatively, the metal ion may combine with a neutral coordinating agent such as an alkylphosphine oxide or a neutral chelating agent such as phenanthroline to form a large cation that can couple with relatively small anions.

Examples of metals in the anionic form include oxyanions such as permanganate, haloanions such as $FeCl_4^-$ or $ZnCl_4^{--}$, and anionic chelates such as the cobalt chelate of nitroso-R-salt. These metal-containing anions couple with a variety of large cations such as substituted oxonium, ammonium, arsonium, and phosphonium ions to give extractable ion pairs. In cases involving oxonium ions, the solvent is usually the source of such ions so that oxygen-containing solvents (ethers, alcohols, ketones, and esters) are necessary for successful extraction.

Organic extractions. Organic extractions may be classified in terms of those solutes which do not involve chemical interactions other than hydrogen bonding and those such as the organic acids and bases which do. In a homologous series, the K_D values would be expected to increase somewhat with increasing molecular weight. Among the organic acids and bases, control of pH in the aqueous phase is of great importance since the ratio of neutral (and therefore extractable) form to charged (water-soluble) form will depend on it. A typical carboxylic acid having a pK of about 5 will be dissociated at a pH (such as 7) at which a phenol (pK about 10) will be neutral.

Methods of extraction. Three methods of liquid-liquid extraction are generally employed in the analytical laboratory: batch, continuous, and countercurrent distribution. The choice of method will depend upon the absolute value of the distribution ratio of the solute of interest, and on its value relative to the distribution ratios of the various interfering materials.

Batch extraction. In this, the simplest and most-used method, the original liquid phase is brought into contact with a given volume of solvent and the two immiscible phases are agitated until equilibrium is attained. The layers are then allowed to settle before sampling. If necessary, the procedure may be repeated after the addition of fresh solvent. The batch extraction process is used to best advantage when the distribution ratio of the solute of interest is large, since a few extractions will effect quantitative separation.

The most commonly used apparatus is a separatory funnel. For extractions into an immiscible solvent of lower specific gravity than water, an ordinary separatory funnel requires a double removal for each batch extraction. To avoid this troublesome operation, special funnels or extractors have been suggested.

Continuous extractions. This method is particularly applicable when the distribution ratio is relatively small, so that a large number of batch extractions would be necessary to effect quantitative separation. Most continuous-extraction devices operate on a principle which involves circulating the extracting solvent by distillation and condensation. The extracting liquid separates out and flows back into a receiving flask, in which it is again evaporated and recycled while the extracted solute remains in the receiving flask. When the solvent cannot be distilled easily, a continuous supply of fresh solvent may be added from a reservoir. These extractions can be run for long periods of time with little or no attention.

Continuous extraction of solid samples is often used in problems involving biological or natural substances and commonly employs the well-known Soxhlet extractor.

In most continuous extractions, the original aqueous phase remains stationary and the solvent is made to flow through this phase. Another type

continuous extraction involves the flow of two liquid phases counter to each other. These continuous countercurrent extractions are used extensively in purifications on an engineering scale.

Countercurrent distribution. Discontinuous countercurrent extractions were first suggested by L. C. Craig as a means of separating several solutes whose distribution ratios were quite similar. This method has received much attention in organic and biochemical separations and is especially attractive when it is not necessary to recover quantitatively all of the solutes but when a "pure" fraction of each solute is needed. Fractionation is accomplished by performing many individual extractions rapidly and in sequence with systematic combination of separated phases from adjacent tubes. A series of separatory funnels may be used, although more elaborate contacting vessels operating by decantation are suggested for large numbers of transfers. The apparatus may be operated automatically. By careful interpretation of the data obtained by analyzing each tube in a countercurrent distribution extraction, a quantitative estimate of the ratio of two solutes whose distribution ratios differ by 10% or less may easily be obtained. *See* COUNTERCURRENT MASS-TRANSFER OPERATION; EQUILIBRIUM, CHEMICAL; SEPARATION (CHEMICAL AND PHYSICAL). [G.H.MO.; H.FR.]

Bibliography: F. R. Bruce et al., *Process Chemistry*, Progress in Nuclear Energy, ser. 3, vol 1, 1956, L. C Craig, *J. Biol. Chem.*, 155(2):519–534, 1944, L. C. Craig, *Anal. Chem.*, 22:1346, 1950; H. Freiser and G. H. Morrison, *Annual Review of Nuclear Science*, vol. 9, 1959; G. H. Morrison and H. Freiser, *Anal. Chem.*, 30:632, 1958; G. H. Morrison and H. Freiser, *Solvent Extraction in Analytical Chemistry*, 1957; E. B. Sandell, *Colorimetric Determination of Traces of Metals*, 2d ed., 1950; R. E. Treybal, *Liquid Extraction*, 1951; A. Weissberger (ed.), *Technique of Organic Chemistry*, vol 3, 2d ed., 1956.

Extraembryonic membranes

Auxiliary structures associated with the embryos of amniote vertebrates. They consist of specialized embryonic tissues that lie outside the body proper of the developing embryo. These structures—the allantois, amnion, chorion (or serosa), and yolk sac—function in the nutrition, respiration, and protection of the embryo *See* ALLANTOIS; AMNION; CHORION; YOLK SAC. [C.B.C.]

Extrapolation

A process in mathematics used to find the value of a function outside its tabulated values. This is done as in interpolation by assuming that over a small range of x the function may be closely approximated by a polynomial or some other readily computed function. *See* INTERPOLATION.

Formulas. Any of the interpolation formulas can be used, therefore, and the desired value of x substituted in them. Thus, for example, if $y = f(x)$ has been tabulated at $x = x_{-m}, x_{-m+1}, \ldots, x_{-1}, x_0$, the Gregory-Newton interpolation formula (1) may be used

$$y = y_0 + u\,\delta y_{-1/2} + \frac{1}{2!}u(u+1)\delta^2 y_{-1} + \frac{1}{3!}u(u+1)(u+2)\delta^3 y_{-3/2} + \cdots + \frac{1}{m!}u(u+1)\cdots(u+m-1)\delta^m y_{-m/2} \tag{1}$$

to determine a polynomial equation passing through the $m + 1$ ordinates $y_{-m}, y_{-m+1}, \ldots, y_{-1}, y_0$. These differences give

$$u = \frac{x - x_0}{h} \tag{2}$$

and the differences $\delta^k y_{-k/2}$, $k = 1, 2, 3, \ldots$, are the same as those used in interpolation.

If $-1 < u < 0$, then $x_{-1} < x < x_0$ and the formula is used to interpolate. On the other hand, substitution of positive values of u permits its use for extrapolation for y beyond y_0, the last value tabulated.

If the function $y = f(x)$ is known, the error introduced by using a polynomial to extrapolate for the value of the function can be expressed by adding to formula (1) a remainder term

$$\frac{1}{(m+1)!}u(u+1) \cdots (u+m)f^{(m+1)}(\xi) \tag{3}$$

where now ξ is any value of x lying between the smallest and the largest of the numbers $x_{-m}, x_{-m+1}, \ldots, x_{-1}, x_0$, and x. This term will be larger for extrapolation, $u > 0$, than for interpolation, $u < 0$, for two reasons. First, since u is positive for extrapolation, the coefficient of $f^{(m+1)}(\xi)$ will be larger. Second, since the range of values permitted ξ is larger, $|f^{(m+1)}(\xi)|$ must be assumed to be larger. It is necessary, in calculating the error, to take the largest absolute value of this $(m+1)$th derivative of $f(x)$ in the above range of ξ. If there is a singularity of $f(x)$ or of its derivatives near the value of x required in the extrapolation, the remainder term in Eq. (3) would indicate that the extrapolation could involve a large error.

If desirable to extrapolate a distance greater than h, the interval of the table, beyond the limits of a table, it may be helpful to proceed by first extending the entries in the table by extrapolation. This may be done by assuming, for instance, that some order of difference remains constant, or by letting u take on positive integral values in Eq. (1). An estimate of the errors introduced by extending the table in this way can be made by attempting to extrapolate for the last few entries in the table from the earlier entries. For values not near a singularity of the function tabulated, these errors should be of about the same size as those introduced in extending the table. Of course, the same number of entries should be added in the two cases.

Having extended the entries in the table by extrapolation, one can look upon the problem of find-

ing y for an intermediate value of x as just the problem of interpolation. If the degree of the interpolating polynomial is large enough, the error of this interpolation can be ignored in comparison with the error in extrapolating for the additional entries.

Example. From the portion of the difference table lying above the line find by extrapolation the logarithms of 1.07, 1.08, and 1.09 and determine a probable limit for the error.

x	$y = \log x$	δy	$\delta^2 y$	$\delta^3 y$
1.00	0.0000 000			
		43 214		
1.01	0.0043 214		−426	
		42 788		8
1.02	0.0086 002		−418	
		42 370		9
1.03	0.0128 372		−409	
		41 961		8
1.04	0.0170 333		−401	
		41 560		7
1.05	0.0211 893		−394	
		41 166		$7+\epsilon$
1.06	0.0253 059		$-387+\epsilon$	
		$40\ 779+\epsilon$		$7+\epsilon$
1.07	$0.0293\ 838+\epsilon\times10^{-7}$		$-380+2\epsilon$	
		$40399+3\epsilon$		$7+\epsilon$
1.08	$0.0334\ 237+4\epsilon\times10^{-7}$		$-373+3\epsilon$	
		$40026+6\epsilon$		
1.09	$0.0374\ 263+10\epsilon\times10^{-7}$			

If it is assumed the third difference stays constant at 7 units in the seventh decimal place, the difference table can be extended, as shown below the line, by working from right to left. For this purpose the ϵs added to the number are ignored.

To determine the maximum error in the extrapolated values it must be recognized that the true third differences would not necessarily be equal to 7×10^{-7}. The maximum error would occur if all these third differences assumed are too high or too low. Therefore the true values of logarithms, rounded off to seven decimal places, will differ from those computed above by less than the ϵs attached to the numbers in the table.

Since a reasonable value for ϵ is ±1, reasonable upper limits for the errors in the extrapolation for the logarithms of 1.07, 1.08, and 1.09 are, respectively, 1, 4, and 10 units in the last decimal place. Comparison of these values with the true values reveals that the errors are actually 0, −1, and −2 units in the last decimal place. It should be clear from this example, however, that the error in extrapolating beyond the limits of a table can be expected to grow rapidly for each new entry added. See GRAPHIC METHODS. [K.S.K.]

Extrasensory perception (ESP)

The alleged phenomenon of perception or awareness of external events in the absence of any sensory stimulation arising from the events.

The question of whether ESP actually exists is a controversial issue among psychologists and other scientists. While a large body of research purporting to support ESP has been accumulated, the majority of scientists are probably skeptical about both the strength of the evidence and the existence of the phenomenon.

Forms of ESP. ESP proper is subdivided into three possible forms. (1) Mental telepathy refers to the case in which one person is aware of an external event through direct sensory perception and another person is not. The former can be called the sender and the latter the receiver, although mental telepathy does not imply a deliberate effort to send or receive on the part of either individual. (2) Clairvoyance refers to the case in which there is only a receiver and an extant event which is not affecting the receiver through any sensory channels. (3) *Precognition differs from clairvoyance* in that the external event has not yet taken place at the time the receiver's perception takes place.

An experiment illustrating all three would be one in which the receiver attempts to specify, in order, the faces of a well-shuffled deck of cards which he cannot see in the following situations: (1) when a sender looks at and concentrates on the face of each card in turn after shuffling has taken place (mental telepathy), (2) when the cards are preshuffled but no one has yet examined the resultant order (clairvoyance), and (3) when the cards are shuffled after the receiver has predicted the order (precognition). Evidence for the presence of ESP depends on whether probability calculations show that favorable results are extremely improbable on the hypothesis of mere lucky guessing. See STATISTICS.

ESP research concentrated first on mental telepathy, second on clairvoyance, and third on precognition. This progression was partly dictated by a search for proper experimental controls for each of the first two ESP situations above. Thus a mental telepathy situation is also a clairvoyance situation because both the event itself and the physiological events produced in the sender by the direct sensory stimulation are extant at the time of the receiver's perception. Similarly a clairvoyance situation is also a precognition situation since the receiver's perception could have been initiated before the event became extant.

For the most part researchers in the field have concluded that results are about as good whether the receiver works with a sender (mental telepathy) or without a sender (clairvoyance), or whether he divines an extant event (clairvoyance) or a future event (precognition). In terms of successive controls this would suggest that clairvoyance can explain mental telepathy and that precognition can explain both mental telepathy and clairvoyance, but workers in the field of ESP have not yet concluded that there is no mental telepathy and tend to regard precognition as a property of clairvoyance—namely its independence of time—rather than as a control for clairvoyance. Another conclusion is that ESP is independent of distance between the receiver and the event; that is, reported results tend to be equally good whether the receiver is only a few feet removed or hundreds of miles removed from the event divined. This conclusion, together with the conclusion that precogni-

tion and clairvoyance work equally well, has helped produce the general conclusion among ESP supporters that it is a nonphysical phenomenon independent of space and time.

Psychokinesis. A relatively recent development leading out of ESP research is psychokinesis (PK), which refers to the alleged ability of an individual to exert a mental influence on physical events in advance of their occurrence. In a typical experiment dice are thrown mechanically while the subject is instructed to will the dice to fall in a particular way (high numbers rather than low numbers, for example). In a sense, PK is a control for precognition since a subject who is striving to be correct in a precognition experiment may be regarded as attempting to predestine the events he is predicting. Reported results of PK experiments have usually been positive, and some of the exponents of this general area of research have grouped ESP and PK under the single heading of psi phenomena (for psychic phenomena) and have labeled the over-all field of study parapsychology.

Elusiveness. In addition to the properties of independence of space and time reported for psi phenomena there is a property which can be called elusiveness. Thus a subject who gives scores significantly better than chance guessing on one occasion may give chance scores on another occasion or may give scores significantly below chance on another occasion. Alternatively it may be found that the subject's attempts to guess the order of cards in a deck are better if one considers his guesses as displaced in each case to the next card, or perhaps to the next-but-one card, a displacement which may be in the forward or in the backward direction.

One general feature of this elusiveness appears to be that performance is better at the outset of a session of trials than at the middle or the end of a session. Another general feature of the elusiveness is that the more informal the testing situation, the better the results; and the more elaborate the scientific precautions surrounding the testing situation, the poorer the results.

Criticism of ESP. Critics of ESP have been numerous, and attacks on the acceptability of the findings have covered such issues as statistical methods, selective reporting of only favorable outcomes, inadvertent sensory cues to the subject, bias in recording errors, and so forth. Some of the objections to earlier work have been taken into account in subsequent work, and the general effect of new precautions has been to reduce scores, although not all the way to the chance level.

Attempts to settle the existing controversy are handicapped by the fact that proponents of ESP state that the [illegible]

[illegible]

[illegible] backward by one or more cards in the sequence, or when over-all results are at the chance level but beginning results are above chance and final results are below chance, or when a U curve is produced by good scores only at the beginning and end of a session, or when some subjects are above chance and other subjects are below chance. Attempts to settle the controversy are also handicapped by the fact that an alleged property of psi phenomena is their tendency to vanish when submitted to stringent tests circumscribed by an atmosphere of scientific rigor.

A final difficulty peculiar to PK is the implication that an experimenter can predestine his data. Thus skeptics who fail to get positive results can be charged with consciously or unconsciously attempting to will their results to come out against psi, that is, to conform to chance guessing. This difficulty in PK research has no counterpart in science proper, in which it is assumed that the motives of the experimenter cannot directly influence the physical events which comprise his data and that even a skeptic will reproduce previous experimental findings. *See* PSYCHOLOGY, PHYSIOLOGICAL AND EXPERIMENTAL. [F.D.S.]

Extrusion

The forcing of solid metal through a suitably shaped orifice under compressive forces. Extrusion is somewhat analogous to squeezing toothpaste through a tube, although some cold extrusion processes more nearly resemble forging, which also deforms metals by application of compressive forces. Most metals can be extruded, although the process may not be economically feasible for high-strength alloys.

Hot extrusion. The most widely used method for producing extruded shapes is the direct, hot extrusion process. In this process, a heated billet of metal is placed in a cylindrical chamber and then compressed by a hydraulically operated ram (Fig. 1). The opposite end of the cylinder contains a die having an orifice of the desired shape; as this die opening is the path of least resistance for the billet under pressure, the metal, in effect, squirts out of the opening as a continous bar having the same cross-sectional shape as the die opening. By using two sets of dies, stepped extrusions can be made. The small section is extruded first to the desired length, the small split die is replaced by the large die, and the large section is then extruded.

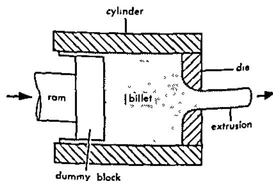


Fig. 1. Schematic representation of direct extrusion process (hot).

Extrusion pressures and speed vary considerably depending upon the size and shape of the section and the mechanical properties of the metal. Some metals, such as magnesium and some aluminum alloys, require slow speeds of a few feet per minute, while others, such as some copper alloys, lead and steel, are extruded at speeds of over 1000 ft/min. The extrusion speed is also somewhat dependent upon the temperature of the alloy. Considerable heat is generated by the process; and if the extrusion speed is high, this heat cannot be dissipated, resulting in a rise in temperature. In some instances the rise in temperature may be sufficient to melt or at least weaken the metal to the point where the frictional stresses at the surface cause cracking.

The flow of metal is not uniform during extrusion of the billet. Because of the restraining effect of the die face and the frictional effects at the cylinder walls, the outer zones of the billet resist deformation and the flow occurs most rapidly at the center of the billet. Eventually, as the billet shortens, the different rates of flow at the center and surface result in the billet's becoming hollow. If the extrusion ram travels further, defects appear at the center of the extruded section. Therefore, a portion of the billet may be left in the cylinder and discarded. To prevent the oxidized surface layers of the billet from getting into the extruded product, the dummy block (or follower plate) in front of the ram is of slightly smaller diameter than that of the cylinder; thus, a thin sleeve of the billet is extruded out of the ram end of the cylinder and is subsequently discarded.

Lubricants are used to minimize friction and protect the die surfaces. Graphite is a common lubricant for nonferrous alloys, whereas for hot extrusion of steel, glass is an excellent lubricant.

Indirect or inverted extrusion was developed to overcome such difficulties as surface friction and entrainment of surface oxide of direct extrusion. In the indirect process the ram is hollow, the die opening being in the dummy block, and the opposite end of the cylinder is closed. As the ram advances, the billet does not move as in the case of direct extrusion, and the metal is extruded backward through the die and the hollow ram. However, the process is not too popular because the hollow ram is weaker, resulting in lower machine capacity; trouble-free operation requires that the extruded product be straight and not hit the inside of the ram.

Tubular shapes are produced by a mandrel of the desired inside shape of the product. The mandrel may be fastened to the dummy block in a direct extrusion press if a pierced billet is used; or in the case of a solid billet, a mandrel must first pierce the billet, after which the main ram advances.

In the production of lead cable sheathing or cored solder wire, the core material (cable or rosin flux) passes through a core tube (or die block) around which is heated lead under pressure. The core material passes from the core tube into a die cavity and the lead is extruded out with it, forming a casing around the cable or rosin core. The process is semicontinous, molten lead being added to a vertical cylinder and pressure applied periodically. The lead is solid when it reaches the core material.

Cold extrusion. The extrusion of cold metal is variously termed cold pressing, cold forging, cold extrusion forging, extrusion pressing, and impact extrusion. The term cold extrusion has become popular in the steel fabrication industry, while impact extrusion is more widely used in the nonferrous field

The original process (identified as impact extrusion) consists of a punch (generally moving at high velocity) striking a blank (or slug) of the metal to be extruded, which has been placed in the cavity of a die. Clearance is left between the punch and die walls; as the punch comes in contact with the blank, the metal has nowhere to go except through the annular opening between punch and die. The punch moves a distance that is controlled by a press setting. This distance determines the base thickness of the finished part. The process is particularly adaptable to the production of thin-walled, tubular-shaped parts having thick bottoms, such as toothpaste tubes.

A process requiring less pressure than backward extrusion is the forward extrusion process, originally called the Hooker process (Fig. 2). A formed blank (usually a thick-walled cup) is placed in a die cavity and struck by a punch having a shoulder or enlarged section a short distance from the end. Upon contact with the blank, the nose or end of the punch starts to push the center of the blank through

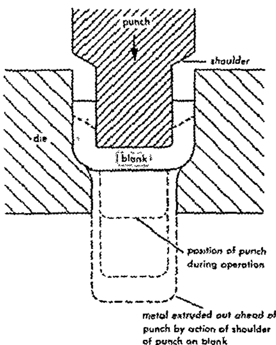


Fig. 2. Principles of forward cold extrusion process (Hooker process).

the die cavity, similar to the action occurring in deep drawing of sheet metal (*see* SHEET METAL FORMING) After the punch has advanced a short distance, the shoulder comes in contact with the top of the thick wall of the blank. The punch shoulder then extrudes the metal through the annular space between the die and the end of the punch. Thus, in forward extrusion, the metal moves in the same direction as the punch, whereas in backward extrusion the metal moves in the opposite direction.

The application of the cold extrusion process to steel was developed in Germany about 1930 but was not released until after World War II. The success of cold extrusion of steel hinged on the discovery of a suitable lubricant and surface treatment. The steel is given a phosphate surface coating, which absorbs and holds the lubricant (soap emulsion, vegetable oils, or dry metal stearates) and prevents seizure between the metal and tools.

Advantages of cold extrusion are higher strength because of severe strain-hardening, good finish and dimensional accuracy, and economy due to fewer operations and minimum of machining required *See* METAL FORMING. [R.L.F.]

Eye

A sense organ found in most vertebrates and in certain groups of invertebrates. The eye is a photoreceptive structure capable of image formation. Simple eyes, or *ocelli*, as well as compound eyes occur in the invertebrates. The ocelli are predominantly organs for light detection whereas the compound eye forms images and in the mollusk *Sepia* is functionally and structurally as complex as the vertebrate eye. *See* PHOTORECEPTION; VISION.

EMBRYOLOGY

In many respects the origin and development of the eye and its parts are similar in all vertebrates from fishes to man. The eyes appear very early as vesicles growing from the side walls of the primitive brain. If their lateral location is not permanent they are eventually, as in man, carried around to the front by the growth of the face. The central nervous system, the brain and spinal cord, first appears as a shield-shaped thickening or neural plate in the surface or ectodermal cells. This plate elongates; its lateral edges rise as folds (Fig. 1*a*), approach each other in the midline, fuse, and form the primitive neural tube. *See* NERVOUS SYSTEM; NEURULATION.

Induction. Experiments on amphibian embryos show that the origin and early development of the neural plate depend upon an inductive influence from the substrate layer of chordo-mesoderm cells. A tonguelike extension of this substrate (Fig. 1*b*) under the forepart of the plate has the capacity to induce, normally, two optic vesicles which grow out when the neural folds come together. If the tonguelike substrate is gradually narrowed by excising more and more of its lateral borders, the two optic vesicles growing later from

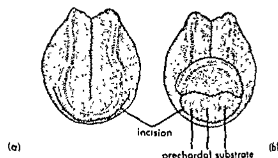


Fig. 1 (a–b) Neural plate stage of amphibian embryo shows incision line and elevation of the anterior neural plate to expose underlying mesoderm (prechordal substrate) which was excised to produce eye abnormalities. (From L. S. Stone and F. L. Dinnean, *Yale J. Biol. and Med.*, 16.37, 1943–1944)

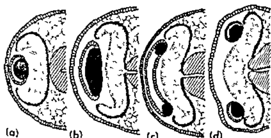


Fig. 2. (a–d) Amphibian embryos showing progressive fusion of eyes and lenses which resulted from cyclopian defects produced by removing more and more of lateral borders of prechordal mesoderm under neural plate. (From L. S. Stone and F. L. Dinnean, *Yale J. Biol. and Med.*, 16.43, 1943–1944)

the brain will in turn show all degrees of fusion as they arise closer and closer together (Fig. 2*a* and *b*). Finally, with the narrowest median strip of eye-inducing substrate left, only a single large median eye (Fig. 2c and *d*), a cyclops, is formed, which is similar to an abnormality occurring in all vertebrates, including man. If the entire substrate is removed, no eyes develop. *See* EMBRYONIC INDUCTION.

Optic vesicle. When the outgrowing optic vesicle touches the surface ectoderm, the latter becomes a thick disk which soon develops also into a vesicle (Fig. 3) that pinches off and transforms into a lens (Fig. 4). Experiments on embryos of lower vertebrates show that a lens will develop from body ectoderm when the optic vesicle comes in contact with it. Therefore, the optic vesicle appears to play a main role in the origin of the lens. However, it is not clear in some vertebrates whether the eye-inducing mesoderm already possesses a feeble specificity for inducing the lens.

As the lens vesicle sinks inward, the optic vesicle invaginates to form an irregular cup to receive it (Fig. 3). The outer covering of the cup becomes the thin pigmented layer of the retina, whereas the invaginated portion thickens to become later the neural or sensory retina. Experiments on amphibian embryos demonstrate that at the primitive optic

cup stage the future functional fields of the retina are already predetermined before the retina differentiates. If the embryonic eye at this stage or later in development is excised and reimplanted in the same place, visuomotor responses later are normal if the eye is normally oriented, but entirely reversed if the graft is oriented upside down (180° rotation). However, if eyes at earlier stages are similarly excised and reimplanted, vision later is always normal regardless of orientation. This definitely proves that the future visual pattern in at least one vertebrate eye is the result of a very early organization taking place in a membrane in which rapid cell proliferation is still progressing

Because the lower margin of the optic cup grows less rapidly than the upper rim, a cleft or fetal *choroid fissure* results, extending along the under-

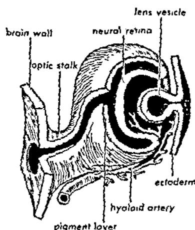


Fig. 3 Schematic representation of optic cup growing from brain wall, lens vesicle from surface ectoderm, and hyaloid artery entering cup through fetal fissure Approximately 5-week human embryo (7.5 mm) (Modified from I. Mann from D. B. Kirby, Surgery of Cataract, Lippincott, 1950)

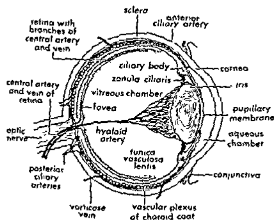


Fig 4. Schematic plan of an eye of a sixth-month human fetus showing relations of structures before the tunica vasculosa lentis and the vascular pupillary membrane degenerate.

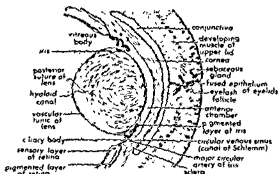


Fig. 5. Vertical section of the anterior part of the eye from a human fetus of about 19 weeks (174 mm). Eyelids fused at this time. (From B. M Patten, Human Embryology, Blakiston–McGraw-Hill, 2d ed., 1953)

surface of the optic stalk (Fig. 3). In mammals, a small blood vessel enters this fissure, which later normally closes around the artery. A failure of the fissure to close produces a defect, a ventral notch in the pupil, called a coloboma. The centrally enclosed artery enters the cup as the hyaloid artery and branches profusely to form a vascular basket, the tunica vasculosa lentis (Fig. 4), about the early developing lens. In the human fetus this reaches maximum development during the third to fifth months and normally degenerates from the sixth to the eighth months, leaving occasionally only a small vestige of the original artery at the time of birth.

Lens. As the hollow lens vesicle in the human embryo pinches off from the surface ectoderm during the fifth to the sixth weeks it sinks into the optic cup (Fig. 3). While the rest of the eye is developing, the lens vesicle becomes filled with a clear crystal body formed by the accumulation of long ribbonlike fibrous cells derived from its inner pole. The outer anterior pole remains as a cap of a single layer of cuboidal cells while the entire surface becomes covered by a thin, highly refractile lens capsule Further growth takes place from a peripheral equatorial band of cells (indicated by streak arching across lens in Fig. 5) which continuously adds layers of new transparent lens fibers even in adult life. Suture lines are produced where the ends of the fibers come together (Fig. 5). Because the lens possesses no direct blood supply it depends upon the surrounding aqueous and vitreous humor environment (Figs. 4 and 5) for carrying on its anaerobic metabolism

Retina. The epithelial walls of the optic cup embracing the lens lay down the early stages of the true retina, the pars optica retinae. The outer pigmented and nonpigmented epithelium at the rim of the optic cup soon proliferate nonnervous epithelium, sometimes known as the pars caeca retinae. A wavy line, the ora serrata, separates the two regions; it is that portion in Fig. 5 where the sensory and pigment layers of the retina join the folds in the ciliary body. Proliferation of the t epithelial layers from the rim of the optic c

tends forward as a thin membrane, the pars iridica retinae, surrounding the pupillary space. Its outer surface is covered by a thin film of mesenchymal cells and blood vessels extending from regions bordering the rim of the cup. This, along with the underlying pigmented epithelium, is the iris (Fig. 5) From the epithelial layer, sphincter and dilator muscle fibers develop, so that the iris diaphragm, by changing the size of the pupil, can regulate the amount of light reaching the back of the eye. However, for a time during fetal life an elaborate network of blood vessels extends over the pupillary space as a membrane. But, like the tunica vasculosa lentis, with which it communicates, this network normally atrophies before birth, although remnants may persist.

Ciliary processes. In the meantime, the nonnervous epithelium near the ora serrata becomes

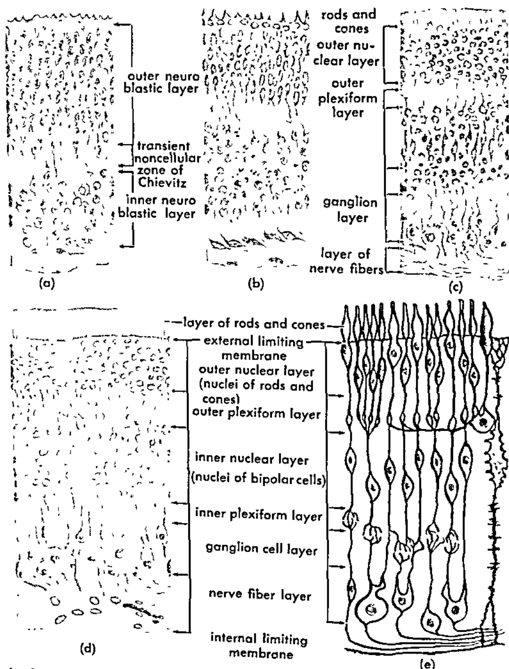


Fig. 6 Stages in histogenesis of neural retina in human eyes. Retina pigment layer not shown. (a) From embryo of about 7 weeks (17 mm). (b) From fetus of about 11 weeks (65 mm). (c) From fetus of about 19 weeks (170 mm). (d) From fetus of 27 weeks (250 mm). *(After I. Mann from B. M. Patten, Human Embryology, 2d ed., Blakiston–McGraw-Hill, 1953)* (e) Schematic plan to show relations of various elements in neural retina. *(From B. M. Patten, Human Embryology, 2d ed., Blakiston–McGraw-Hill, 1953)*

thrown into folds or ciliary processes. This part of the epithelium is known as the pars ciliaris retinae. The adjacent outer layer of mesenchyme differentiates into a thickened muscular layer and, together with the pars ciliaris retinae, it forms a ring known as the ciliary body (Figs. 4 and 5).

The lens is suspended by a circular ligament attached to it behind the pupil, the zonula ciliaris (Fig. 4). This develops in human fetuses during the third and fourth months as a circular membrane of delicate fibers growing from the ciliary processes which are attached to a thick pigmented muscular ring, the ciliary body, at the base of the iris. To focus on distant or near objects, some vertebrate eyes are provided with a special muscle arrangement to shift the lens backward and forward, as is done in a camera. In some vertebrates, such as the seal, there is a pinpoint pupil which enhances the sharpness of the visual image. In some animals the shape of the eye and retina becomes modified during development to provide another method for focus of objects at various distances. In other eyes, including the human, the biconvex lens is elastic enough so that when the suspensory ligament, zonula ciliaris, is under tension, the lens flattens for distant vision. Relaxation under the action of the ciliary muscle allows the lens to return to a more convex condition for focusing on closer objects.

Retinal cell layers. The inner neural or sensory part of the retina in all vertebrate eyes has three prominent cell layers, separated by inner and outer plexiform layers of fibers arising from the cells (Fig. 6*d* and *e*). Minute studies show the various zones indicated in Fig. 6*e* and nonnervous supporting elements, such as Müller cells. Light passes down through the retina to the photoreceptors, the rod- and cone-shaped elements lying on the basal pigment cells. The rods contain a photolabile chemical substance, visual purple or rhodopsin, which bleaches rapidly in very low intensities of light. The cones, containing very little if any rhodopsin, are stimulated by high intensities of light; they are also receptors for color vision. Processes from the rods and cones communicate (synapse) with fibers of the bipolar cells. Processes from the intermediate bipolar cells relay sensory stimuli from the rods and cones to fibers of ganglion cells. Nerve fibers from the ganglion cells sweep over the inner surface of the neural retina and under the internal limiting membrane to a point of exit, and collectively they form the optic nerve which relays the impulses to the proper brain centers. The head of the optic nerve is the blind spot in the retina because it contains fibers but none of the retinal cells. *See* COLOR VISION.

Human [illegible]

[illegible] seventh month (Fig. 6*d*) the retina has almost the same appearance as that found in the adult eye. The inner ganglion cells have already sent their fibers through the optic nerve to the brain. The rods and cones are the last to complete their differentiation. The eye is now potentially functional. The whole eye increases in volume even beyond the time of birth, but the retina keeps pace chiefly by spreading and thinning its cellular layers.

A central depression, called the macula or fovea centralis, appears in the retina by the seventh month as the upper layer of cells gradually spreads aside, leaving exposed a small thin spot of cells at the base composed only of cones which require more intense light. This is the spot of sharpest visual acuity where the cones send their stimuli directly to surrounding ganglion cells. It is highly developed in birds and some reptiles. The completion of the fovea in man is apparently not accomplished until about the fourth month after birth. This seems to account for the observation that visual acuity is not well developed in the newly born infant. However, some infants as early as 75 hours after birth can at least follow moving objects with eye and head movements.

Retina in nocturnal animals. Some animals with nocturnal habits develop retinas possessing only rods, which are sensitive to dim light. Many deep-sea fishes develop retinas, as well as large lenses and pupils, which are modified to cope with the dim light of their environment. A few retinal membranes, as in some bats, develop numerous conical elevations through the pressure of fingerlike projections of blood vessels instead of all the cell layers following the spherical wall of the eye as they do in most eyes. This brings the photoreceptor cells over a wavy surface at various distances from the lens, thus providing a mechanism for accommodation, even though vision may be slightly blurred by the cells not in focus. Tubular eyes of some deep-sea fishes develop accessory retinal patches at different distances from the lens, to provide sharp vision for various distances. The shape of the eye of the horse has been described as one which presents a slanting retinal membrane. It has been suggested that tilting such an eye up or down will provide a retinal image over a considerable distance.

In some animals, such as the blind fishes and salamanders found in caves, the eye may proceed in development and then at various periods undergo extensive degeneration. In many moles, for example, degeneration finally reduces the eye to a mere vestigial structure lying sealed under the skin.

Blood vessels. In many vertebrate eyes, blood vessels are prominent in or on the surface of the retina. In the human fetal eye, at 4 months, two branches develop at the base of the hyaloid artery as it enters the optic nerve head and for the next 4 months spread branches through all parts of the retina (Fig. 4), while the temporary tunica vasculosa lentis of the hyaloid artery degenerates back to the head of the optic nerve. These retinal vessels are branches of the central artery and vein in the optic nerve and can be seen in the adult eye through an ophthalmoscope.

A prominent pigmented vascular cone, called the pecten, develops as a peculiar feature in the eyes of birds. It extends into the vitreous chamber of the eye from the head of the optic nerve. A somewhat similar structure is found in reptile eyes. There has been much speculation regarding the possible function of this vascular shadow-casting structure.

Under the vertebrate retina lies a fine plexus of pigmented blood vessels, called the choroid coat (Fig. 4). This vascular network develops from mesenchymal cells surrounding the optic cup and appears in the embryonic human eye shortly after the first month in development. Pigment soon covers these vessels. The network grows forward and connects with those that later develop in the ciliary body and in the iris (Fig. 4). This pigmented vascular coat lies next to the pigment layer in the retina, except in those animals such as the cat, in [illegible] highly iri-

brane which reflects light at night

Sclera and cornea. Most of the outer wall of the eyeball consists of a thick opaque fibrous layer, the sclera (Figs. 4 and 5), a portion of which is seen as the white part of the eye. Continuous with it in front of the iris and pupil is a relatively small, clear disk, the cornea (Figs. 4 and 5). Early in development the optic cup becomes covered by a layer of opaque mesodermal cells. After the lens has been pinched off, the surface ectoderm with the underlying mesoderm in front of the cup becomes transformed into a transparent corneal membrane. It consists of an outer, thicker, stratified epithelial covering, arising from the surface ectoderm. A thick layer of fibers and a thin, inner single layer of endothelial cells are derived from the mesoderm. The various layers are well defined by the second month in human embryos (Fig. 5). The time when the cornea becomes transparent varies in different vertebrates. The precision with which the cornea is molded is important for obtaining good vision. Sight is often restored to an eye blinded by a scarred, opaque cornea by replacing the latter with a transplant of clear cornea taken from another eye. The donor eye is either available at the time of an operation or has been kept in storage for a number of hours.

A gradual condensation of the mesoderm about the rest of the optic cup transforms it into the tough fibrous sclera (Fig. 4), which is well developed by the fifth month in the human fetus. It extends from the margin of the cornea and joins the covering over the optic nerve (Fig. 4). A thin plate of cartilage develops in the sclera of some vertebrate eyes.

Aqueous chamber. During development the spaces between the cornea, iris, and lens (Fig. 4) become filled with a watery fluid, the aqueous humor. This is secreted by cells in the ciliary region and drained into the venous channels by a system of sinuses and the canal of Schlemm (Fig. 5), that begins to appear along the corneoscleral junction about the third month in human fetal eyes.

The large inner space, the vitreous chamber (Fig. 4), behind the lens and bordered by the retina, is filled with a gelatinous transparent material called the vitreous body. Fluid appears early within the optic cup and as the cup enlarges mesenchymal cells from around the lens invade it. In a vertebrate eye such as that of the human, the vitreous chamber is partly occupied for the first few months by the vascular tunic (Fig. 4) around the lens. With its disappearance the gelatinous vitreous body continues to keep pace with the growth of the eye, and along with the aqueous humor, maintains the intraocular pressure.

Eyelid. In those animals which possess eyelids, the lids begin to form some time after the eye has begun to develop. Experiments on embryos of lower vertebrates show that the eye induces the lids to form from surrounding tissues. In human embryos eyelids appear during the seventh week as folds in the skin above and below the eye. They come together rapidly and fuse about the ninth week. Fusion takes place (Fig. 5) only in the surface epithelial cells above a line where the eyelashes and associated small glands later develop. Between the sixth and seventh months of fetal life the eyelids open and remain separated permanently. However, in many mammals, such as the cat and dog, the lids do not open until sometime after birth. In snakes, the eyelids fuse, remain closed, and as modified lids become a transparent membrane protecting the cornea.

Conjunctiva. After the lids close, about the seventh week in human embryos, a highly vascularized, moist, saclike membrane, the conjunctiva, lines the lids and spreads out over the white sclera of the eyeball (Fig. 5), only as far as the edge of the clear cornea (Fig. 4). At this time the lacrimal or tear glands develop from epithelium of the conjunctiva. They secrete the watery fluid that keeps the surface of the eye moist and cleaned. This fluid drains from the inner corner of the eyelids into the nose, through a nasolacrimal duct, which is laid down chiefly between the sixth and eighth weeks by the sinking inward of a surface groove during the formation of the face. *See* LACRIMAL GLAND.

In some vertebrate eyes the conjunctival sac develops a vertical fold, the nictitating membrane, at the inner nasal corner of the lid opening. It is transparent, and being provided with a muscle, it sweeps the cornea clean. It is easily observed in the eyes of cats and dogs. In birds it can keep the cornea moist while in flight. In lizards it protects the eye from injuries in its environment, and in the horse and other grazing animals it no doubt has a similar function. In the eye of the tadpole, the nictitating membrane develops early. It is continuous with the skin and because it covers the cornea it is called the spectacle. After metamorphosis, when lids develop, it is replaced by a broad transparent

retractile fold of the inner lining of the lower lid which develops to sweep upward over the eye as the nictitating membrane.

Ocular muscles. The presence of the eye is essential for the development of the ocular muscles, which become attached to the scleral wall, and are necessary for movement of the eye in all directions. As early as the fifth week in human embryos they can be seen condensing from the mesoderm surrounding the optic cup. Experiments on amphibian embryos have shown that if the eye-forming area is removed very early, no muscles develop from the mesoderm which normally surrounds the eye. The size of the muscles is also determined by the size of the eye. Therefore, from these and other studies of the vertebrate eye it is evident that the early optic vesicle and cup induce other specific structures to develop within the environment. [L.S ST]

ANATOMY AND HISTOLOGY

Vertebrate eyes are paired sense organs in which visual images are focused on a sheet of light receptors, where photochemical reactions initiate neural impulses that are relayed, via the optic nerve, to the brain.

The anterior segment gathers and transmits incident light and inverts and focuses images. It is comprised of the cornea, eyelids, iris, lens, ciliary body, sclera, choroid, and retina.

Cornea. The cornea, a circular, transparent window, has five layers (Fig. 7). The outermost layer, the stratified squamous epithelium, is thick in fish and superficially keratinized in aquatic mammals and anteaters. It rests on Bowman's membrane, an acellular lamella, which is either lacking or difficult to distinguish in fish, marsupials, and many mammals. The middle layer of the cornea is composed of scattered cells of collagenous fibers arranged in lamellae and of ground substance containing large amounts of mucopolysaccharide. It rests upon the acellular Descemet's membrane, which is in turn underlaid by a single layer of cells, the endothelium, which lines the front of the anterior chamber (Fig. 8).

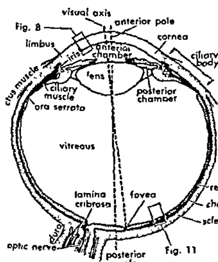


Fig 7. Axial section of the human eye.

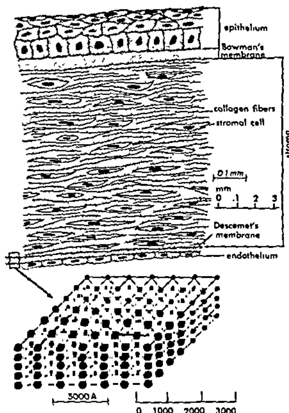


Fig. 8. Transverse section of cornea. Inset shows structure of Descemet's membrane as revealed by the electron microscope (Jakus)

In terrestrial forms, the cornea is the strongest refractor in the eye because its refraction is not offset by a watery environment. In birds and mammals, this property is enhanced by a greater curvature of the cornea than of the remainder of the eye, a trait shared by the skates. Except in some goldfish, the cornea is avascular. It is larger in diameter and has more light-gathering power in nocturnal and deep-sea animals.

Eyelids. The eyelids and nictitating membrane are in front of the eye and protect the corneas of terrestrial animals from abrasion, dehydration, and light of great intensity. These structures have their counterparts in immobile ridges of skin which surround the eyes of many cartilaginous and bony fishes. True eyelids and nictitans are rudimentary in the urodeles, absent in snakes and most nocturnal reptiles, and tend to be absent in some of the higher mammals. The chimpanzee alone among the primates possesses a nictitating membrane. Among the amphibians and reptiles the lower eyelid is movable. Birds can move either the lower lid alone

or both lids. With few exceptions the upper lid of mammals is the more active. The snakes and some lizards have fused and transparent lids which form a protective covering (brille) over the cornea. Some lizards and turtles have a transparent lower lid (spectacle) which can be drawn over the cornea.

Iris. The iris is a diaphragm perforated centrally by an adjustable pupil, which regulates the amount of light traversing the underlying lens. It has five layers: the simple squamous anterior epithelium, the anterior limiting layer of connective tissue; the stroma, containing connective tissue, pigment cells, vessels, nerves, and, near its pupillary margin, the sphincter muscle responsible for pupil constriction; the posterior limiting layer, containing radially disposed contractile cell processes in the posterior epithelium (dilator of the pupil); and the pigmented, stratified posterior epithelium. The iris muscles are striated in reptiles and birds and smooth in all other forms

Although the iris is rudimentary in some deep-sea fish, it is present in all vertebrates. Its color is determined by the presence of silvery guanin (fish, some reptiles), red or yellow carotenoid droplets (some birds), or melanin (most forms). Pupil shape may be irregular (many fish and mammals), a vertical slit (some fish, some amphibians, reptiles, some mammals), a horizontal slit (some fish, some amphibians, some mammals), or circular (birds, some mammals).

Protuberances from the dorsal or ventral iris margin (called corpora nigra in mammals, and opercula in lower forms) are present in representatives of each class except the birds. They facilitate complete closure of some pupils in bright light.

Except in lampreys, sharks, sturgeons, and lungfish, a pectinate or annular ligament of cells and fibers stretches between the outer edges of the iris and cornea This meshwork filters the aqueous humor as it passes from the anterior chamber into the drainage canal of Schlemm.

Lens. This structure is present in all but a few rudimentary eyes and aids in focusing the visual image. Lenses tend to be spherical in aquatic and nocturnal species, and ovoid or of other shapes in diurnal terrestrial forms. The lens is the strongest refractor in eyes of aquatic animals. It is composed of an acellular capsule which is underlaid anteriorly by the lens epithelium and which encloses a core of regularly arranged, elongated cells, the lens fibers In turtles, lizards, and birds these fibers are thickened at the equator to form a pad, the *Ringwulst*, which facilitates changes in lens shape during focusing. The fibers meet anteriorly under the epithelium, and posteriorly under the capsule, in suture lines which have a Y-shaped pattern in mammals, are straight and vertical in bony fish, sharks, frogs, and some birds, and are star-shaped in some fish, a few amphibians, reptiles, and most birds (Fig. 9).

Objects at different distances are sharply focused by moving the lens or by changing its shape

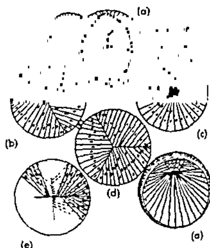


Fig. 9. Anterior lens suture patterns. (a) Sharks, some teleosts, rabbits, some amphibians, and some birds (b) Most mammals. (c) Posterior surface of (b). (d) Most birds, some teleosts, some amphibians, some reptiles (e) Frogs (f) Diagonal section through (c). (*Redrawn from J H Prince, Comparative Anatomy of the Eye, Charles C Thomas, 1956*)

(accommodation). Lower vertebrates utilize lens movement. In lampreys and bony fish the lens is at rest for near vision and is moved backward for distant vision. The lampreys move the lens by depressing the overlying cornea with the cornealis muscle. The bony fish use the retractor lentis muscle which attaches to the lens and to a vascular body (falciform process) in the eye wall, or directly to the sclera. The eyes of sharks, amphibians, and snakes are at rest in distant vision and the lens is advanced to focus on near objects. For this purpose the sharks and amphibians employ protractor lentis muscles which attach to the lens or zonular fibers and to the iris or cornea. In land snakes, contraction of the iris musculature places the vitreous under pressure and moves the lens forward. In birds, most reptiles, and mammals the eyes are at rest in distant vision, and lens shape is changed during accommodation. Birds and mammals effect such changes by contraction of the ciliary muscles, but the sea snakes, lizards, turtles, and some tortoises use the compressive power of the iris sphincter muscle (Fig. 10).

Ciliary body. The ciliary body is a ring of tissue which lies just anterior to the retinal margin and gives rise to the iris and to the suspensory ligament of the lens. Its core is connective tissue and contains the ciliary muscles which change lens shape during accommodation. These muscles are poorly developed in fish and amphibians and well developed in reptiles (except snakes), birds, and mammals. The epithelium of the ciliary body has two layers which are continuations of the pigmented epithelium and neural retina. It secretes aqueous humor. In addition it serves to anchor the zonular fibers (collectively the suspensory ligament of the lens), which attach it to the lens equator. In rep-

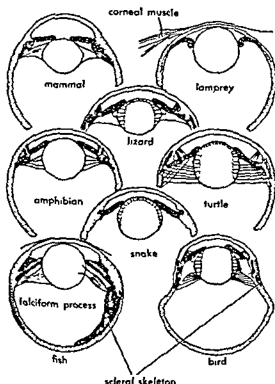


Fig. 10. Mechanisms for accommodation used by vertebrates.

tiles (except snakes), birds, and mammals its inner surface is thrown into radial folds, the ciliary processes, which surround the lens equator.

The posterior portion of the eye forms a suitable screen for the visual image and a mechanism for generating neural impulses from light. The posterior segment of the eye is invested with three concentric coats, the sclera, choroid, and retina, described in the following sections.

Sclera. The sclera, the outside coat, serves to maintain eye shape and to attach the muscles that move the eye. The shape of the posterior segment varies from tubular in some deep-sea fish and a few birds (owls, some eagles, hawks), to oblate in many fish, amphibians, reptiles, and birds, to spherical in snakes and mammals. All the major submammalian groups, with the exception of lampreys and snakes, possess bony or cartilaginous plates in the eye wall (sclera). This ocular skeleton permits the assumption of shapes other than spherical and minimizes distortion of the eye during accommodation. The number of scleral ossicles varies from 2 per eye (fish) to a maximum of 18 per eye (birds). The mammals, like the lampreys and snakes, have a fibrous sclera.

The extrinsic eye muscles are the internal and lateral recti (horizontal movement); the superior and inferior recti (vertical movement); the superior and inferior obliques (rotation); and the retractor and levator bulbi muscles for retracting and protracting the eye. The last two muscles appear in the amphibians, and with the exception of birds and primates, are present in most forms in which they are also responsible for lid movements. The basic muscle plan varies only in details in the vertebrate series.

Choroid. The choroid coat is a highly vascular layer which lines the sclera. Its thickness and vascularity are inversely related to the degree of vascularization of the neural retina. The extreme case is seen in the eel in which a highly vascularized retina exists without any choroidal tissue.

In some species the choroid contains pigment cells (melanocytes) throughout its extent, and it functions to absorb light which has traversed the retina, thus preventing back scatter. In other forms, the choroid, or a portion of it, is modified by the presence of a tapetum, a reflecting layer situated behind the neural retina. Choroidal melanin is absent in tapetal regions. The tapetum serves to pass unabsorbed light back through the retina, and thereby increases visual sensitivity, at the expense of visual acuity, in many aquatic and nocturnal animals. It is responsible for "eyeshine." The tapetum is absent from amphibians, most reptiles, and birds. In fish it consists of guanin crystals deposited in the retinal epithelium, or, more rarely, in the choroid. The tapeta of mammals are fibrous layers (in most ungulates) or cellular layers (in many carnivores) in the choroid, usually in the upper half of the eye.

Blood vessels are brought deeper into the eye in a variety of ways. The mammals have blood vessels which radiate from the head of the optic nerve in the innermost layers of the retina. The submammalian orders of vertebrates rarely have retinal vessels. In some of these lower forms the supply takes the form of a vascular protrusion from the head of the optic nerve into the vitreous body. In birds it is pleated and is called a pecten. In the lizards and a few snakes it is cone-shaped and is called the conus papillaris. In some fish it is a flattened structure called the falciform process, the anterior end of which (campanule of Haller) attaches the retractor lentis muscle. In most fishes, snakes, and anuran amphibians, a vascular net occurs at the interface between the vitreous body and the neural retina. Some forms such as the urodele amphibians and some fish have no special provision for deep vessels in the eye (Fig. 11).

Retina. The retina is composed of a pigmented epithelium and a neural layer and encloses the jellylike vitreous body.

The outermost layer of the retina lining the choroid is a single layer of flattened, hexagonal cells with processes which extend between the rods and cones. In most species these cells contain brown pigment granules (melanin). This pigment is most abundant in the bony fish and birds. It is lacking wherever a tapetum occurs and in albino animals. In general the inframammalian species show migration of the melanin granules into the processes during light adaptation to increase acuity, and a retreat of the granules during dark adaptation to increase sensitivity.

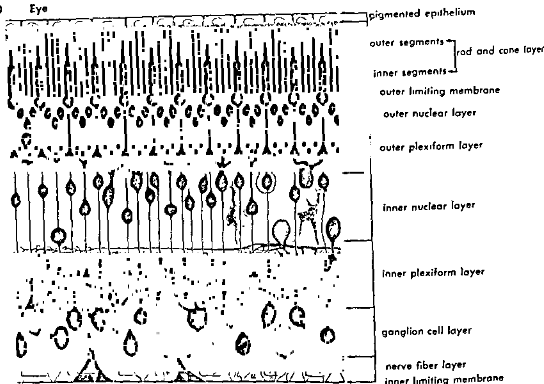


Fig 11 Schematic representation of the retina in cross section. (*Polyak*)

The neural retina consists of an outer layer of light receptors and two additional layers of neurons. The outermost (bacillary) layer contains light-sensitive rods and cones. The cones are thinner and less sensitive to light than rods and participate in daylight (photopic) and color vision Rods operate in dim light (scotopic vision). Sharks, rays, some crocodiles, most snakes, and the nocturnal species of all classes have predominantly rod retinas The diurnal forms of fish, amphibians, and mammals have mixed or predominantly cone retinas. In many diurnal forms of all classes a small region of the retina (the macula) is packed with thin receptors (usually cones), is specialized for fine vision, and serves as a point of visual fixation. A small retinal depression, which increases visual acuity, is associated with the macula in a few bony fish, most lizards, most birds, and some mammals It is centrally located in the retinas of fish, most lizards, turtles, tortoises, most birds, and mammals; laterally placed in the fast-striking tree snakes; and present at both locations in the fast-flying hunting birds

Rods and cones have outer and inner segments. Each outer segment is composed of a stack of lamellae, a longitudinal bundle of fibrils, and an enveloping membrane; this segment mediates the photochemical reaction which triggers the visual nerve impulses. The inner segment contains a refractile ellipsoid body and a contractile myoid re- [illegible] ians, and birds shorten in darkness and elongate in light. This assures proper positioning of the receptors for different light conditions. A paraboloid structure of uncertain function is present in the inner segments of the rods of a few species and the cones of all groups except the snakes, bony fishes, and mammals. There are refractile "oil" droplets at the outer ends of the inner segments of the cones of sturgeons, amphibians, turtles, lizards, birds, platypuses, and some marsupials. These droplets are variously and brightly colored by carotenoid pigments in turtles and birds and, as color filters, may participate in color vision in these forms.

The remainder of the retina is concerned with integrating the visual impulses and transmitting the result to the brain. The outer limiting membrane is pierced by the rods and cones and underlain by the outer nuclear layer containing the cell bodies of the rods and cones. These synapse, in the outer plexiform layer, with the bipolar cells. The bipolar cells and the horizontal and amacrine cells which are concerned with spreading neural excitation occupy the inner nuclear layer. The bipolar cells synapse in the inner plexiform layer with the dendrites of the ganglion cell layer. The single axons of the ganglion cells form the nerve fiber layer as they course toward the optic nerve on their way to the brain. The retina is bounded internally by the internal limiting membrane. *See* SPECIALIZED TISSUE; *see also* BLINDNESS. [A.J.CO.]

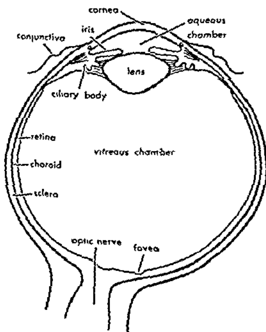
Bibliography: S. R. Detwiler, *Vertebrate Photoreceptors*, 1943; I. C. Mann, *Development of the Human Eye*, 2d ed., 1950; I. C. Michaelson, *Retinal Circulation in Man and Animals*, 1954; W. von Möllendorff, *Handbuch der Mikroskopischen Ana-*

tomie des Menschen, vols. 3–5, 1957; B. M. Patten, *Human Embryology*, 2d ed., 1953; J. H. Prince, *Comparative Anatomy of the Eye*, 1956; L. S. Stone, Return of vision and functional polarization in the retinae of transplanted eyes, *Trans. Ophthalmol. Soc. United Kingdom*, vol. 67, 1947; L. S. Stone and F. L. Dinnean, Lens induction in the salamander (*Amblystoma punctatum*) with special reference to conditions in experimentally produced cyclopia, *Yale J. Biol. and Med.*, 16(1):31–42, 1943; G. L. Walls, *The Vertebrate Eye and its Adaptive Radiation*, Cranbrook Inst. Sci. Bull. 19, 1942; E. Wolff, *The Anatomy of the Eye and Orbit*, 4th ed., 1954.

Eye disorders

The remarkable efficiency of the eye as an organ of vision results from its highly specialized structure and the complex coordination of its component parts which are to a large degree automatic, and each of which is vital to the process of normal vision (see illustration). Damage to any of these structures can result in visual impairment.

Conjunctiva. The conjunctiva (white of the eye) is subject, because of its exposed position, to mechanical and chemical injury and infection. Acute conjunctivitis (pink-eye) may be caused by a variety of bacteria. It is contagious and may occur either in epidemics or in isolated cases following exposure to irritants (smoke, dust, or fumes). It may also accompany other diseases such as measles or scarlet fever. The most common bacterial offenders are staphylococci, streptococci, and pneumococci. Conjunctivitis of the newborn, an infection acquired at birth by passage through an infected birth canal, has all but disappeared because of routine treatment of the eyes of newborn babies with antibiotics or 1% silver nitrate solution. Trachoma is a specific viral disease of the conjunctiva. It is world-wide in distribution, is uncommon in the United States, and is most prevalent in the Middle and Far East where it is a major cause of blindness. Antibiotics are effective in its treatment. Another form of conjunctivitis, more common but less serious than trachoma, is caused by a virus of genital origin and is known as inclusion blennorrhea when it occurs in the infant and as swimming pool conjunctivitis in the adult. Disease of the conjunctiva is doubly important in that it may spread to the transparent cornea and to the delicate inner structures of the eye. See INCLUSION BLENNORRHEA; TRACHOMA.



Normal human eye.

Cornea, iris, ciliary body, and lens. The cornea, like the conjunctiva, is subject to mechanical, chemical, and irradiation injury and to infection. See KERATITIS.

Because of their proximity, both the iris and the ciliary body are usually involved in any disease process that affects one or the other. Such a combination is known as iridocyclitis and is usually secondary to disease elsewhere in the eye or in the body.

The only important disease process affecting the lens is cataract (opacity within the lens). See CATARACT.

Retina. Disease of the retina is usually secondary to disease occurring elsewhere in the body and may be classified into three major groups: (1) infections, which may spread to involve the retina from nearby structures (cornea, conjunctiva, nasal cavities, or skull) or from other organs of the body (kidney, heart, or lungs) via the blood stream; (2) detachment of the retina from the inner surface of the eyeball, which is most common in myopic (nearsighted) eyes and may be caused by injury to the eye, hemorrhage, tumors, scarring, or may occur without known cause; (3) vascular disease of the eye, a heterogeneous group, may result from the formation of a blood clot in the retinal arteries or veins, or may follow hemorrhage and scarring from damaged blood vessels occurring in generalized diseases such as arteriosclerosis, diabetes, and high blood pressure. See RETINITIS.

Optic nerve. The optic nerve, because of its proximity to both eye and brain, may become inflamed as a result of spread of infection from either organ. It is also sensitive to a number of chemicals (lead, thallium, carbon tetrachloride, methyl alcohol, benzene, and benzene derivatives), exposure to which may cause inflammation (optic neuritis). If damage is severe, optic atrophy and blindness follow.

Other disorders. Tumors of the eye are relatively rare. Although more than 50 types of benign and malignant tumors have been described, only 2 of the latter are of major importance. The most common of these, malignant melanoma, occurs in middle and later life; the other, retinoblastoma, is slightly less common and usually appears before the age of 5 years.

Retrolental fibroplasia, a disease of the premature infant, is characterized by formation of scar tissue in the vitreous substance to a degree resulting in permanent blindness. Most, but not all, cases are related to prolonged treatment with high concentrations of oxygen. Limitation of oxygen therapy of premature infants has led to considerable reduction in this disease

Sympathetic ophthalmia is a serious progressive inflammatory condition involving a normal eye following, by a period of usually several weeks, severe injury of its mate. It is thought to be caused by an allergic reaction to displaced choroidal pigment in the injured eye

Glaucoma, a disease of middle and late life, is characterized by increase in the fluid pressure within the eye. See GLAUCOMA.

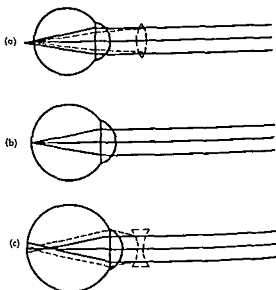
Refractive errors (ametropia) may be any one or a combination of four types Three are abnormal whereas one, presbyopia, is a normal consequence of the aging process Hyperopia (farsightedness), the most common refractive abnormality, is present if the eyeball is too short or if the refractive power of the lens and cornea is too weak. Myopia (nearsightedness) results if the eyeball is too long or if the refractive power is too strong. Astigmatism is present if the refractive power of the eye is not radially symmetrical Presbyopia, a diminished ability to focus on near objects, is caused by the gradual loss of elasticity of the lens with age. See EYE. [W.R.AD.]

Eye glasses

A general term for optical devices containing corrective lenses for defects in vision or for special purposes. Common visual defects are errors of refraction, which is the bending of light rays so that a sharp retinal image is produced. This deflection is produced when light passes from one material to another of different optical density, as from air to water. Besides optical density, the curvature of the surfaces of the material causes either spreading or focusing of light rays, in accordance with the laws of optics. See REFRACTION OF WAVES; VISION.

A normal, or emmetropic, eye will focus light rays from a distance on the retina by means of complementary deflections of the cornea, crystalline lens, and fluid of the eye. Distant light rays, reflected from an object more than 25 ft from the eye, are considered to be parallel. As an object approaches the eye, however, its reflected rays tend to diverge so that, unless some correction is made, these rays will focus behind the retina and vision will be blurred. Such correction is known as accommodation and is achieved through alteration of the lens curvature by action of the ciliary muscle so that a retinal focus is obtained. In infants and older persons, failure of accommodation causes blurred near vision but does not affect far vision.

Presbyopia is the progressive loss of ability to accommodate to near vision as a result of aging processes. It usually becomes apparent in the 40–45 age group and may be corrected by convex



Course of light rays and formation of retinal image in the eye and mechanism of correction by appropriate lenses (a) Hyperopia. (b) Emmetropia (normal). (c) Myopia. (From B A. Houssay et al., *Human Physiology*, 2d ed., McGraw-Hill, 1955)

lenses, usually worn only for close work or reading Bi-, tri-, or tetrafocal lenses are used under different circumstances to permit rapid shifting from one visual distance to another in failures of accommodation.

Few eyes are completely normal in their refractive ability. Defects in curvatures, densities, and position of the eye structures, as well as variations in the length of the eyeball, or visual axis, are called ametropias. Most are hereditary or developmental defects.

Hyperopia, or farsightedness, results from too short an eyeball so that unaccommodated rays focus behind the retina. Moderate defects may be corrected by accommodation, particularly in young people, but the increased accommodation required for close vision often leads to eyestrain. Convex lenses are used to gain proper focus.

Myopia, or nearsightedness, results from too long a visual axis so that the focus falls in front of the retina. The ciliary muscles cannot make the lens of the eyeball less curved. Concave lenses are used to permit focusing of far vision; near vision is often better than that of the normal eye since little accommodation is necessary, less eyestrain is produced with close work, and the retinal image is somewhat larger than in the normal eye.

Astigmatism is a common condition in which the vertical and horizontal meridians of the eye have different curvatures, thereby causing different points of focus. Cylindrical lenses are ground to correct the defect in the abnormal meridian. Most astigmatics also have other defects so that compound lenses are required.

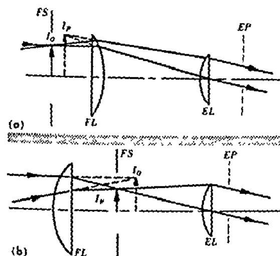
Other lens systems are commonly devised for persons who have unequal vision and to correct for

abnormal visual axes and defects in ocular movements caused by muscle imbalance.

Special forms of eye glasses are made to absorb portions of the light spectrum, such as certain colors and ultraviolet and infrared rays. Aviators, welders, radiologists, furnace workers, and others require lenses of certain physical or optical characteristics. [E.G.ST.]

Eyepiece

A lens or optical system which offers to the eye, at a suitable viewing distance, the real image originating from another system (the objective) as a virtual image (*see* IMAGE, OPTICAL). In modern instruments, most eyepieces (also called oculars) are not independently corrected for all errors but are designed to balance out certain residual aberrations of the objective as, for instance, chromatic difference of magnification. *See* ABERRATION, OPTICAL.



(a) Ramsden eyepiece and (b) Huygens eyepiece. *FL*, field lens; *EL*, eye lens; *FS*, field stop; *EP*, exit pupil or eye point; I_O, image (real or virtual) formed by the preceding system; I_F, image (virtual or real) formed by the preceding system and the field lens.

The Ramsden eyepiece consists of two planoconvex lenses, the field lens and the eye lens, with their plane sides out. Both of these lenses have the same power and focal length; their separation is equal to their common focal length. The field lens is near the image formed by the objective. The Ramsden acts as a field stop. (For a discussion of stops, *see* OPTICS, GEOMETRICAL.) The Kellner eyepiece is a Ramsden eyepiece with an achromatic eye lens. *See* LENS, OPTICAL.

The Huygens eyepiece also consists of two planoconvex lenses, but the plane sides of both lenses face the eye. The focal length of the field lens is in general three times that of the eye lens, and the separation is twice the focal length of the eye lens. The Huygens eyepiece is usually corrected for astigmatism and distortion. A type of Huygens eyepiece in which the eye lens is achromatized to compensate for the color errors of the objective is called a compensating eyepiece. The field lens may also be achromatized.

Huygens and Ramsden eyepieces are in general used for low magnifying powers. For higher powers, eyepieces assume more complex forms to correct astigmatism, field curvature, and distortion for larger field angles.

The orthoscopic eyepiece consists of a single lens, made up of three cemented elements, to which a planoconvex lens is added.

Complex types of oculars enable corrections to be made over a field angle of 70° or more on the image side. They are used especially in military instruments. [M.H.]

Facies (geology)

Any observable attribute of rocks, such as over-all appearance, composition, or conditions of formation, and changes that may occur in these attributes over a geographic area. The word facies has several meanings in common geological usage. It may refer to bodies of specific rock content or combinations of rock content, such as redbed facies and black shale facies; it may refer to some areally restricted and mappable part of a stratigraphic rock body; or it may be used as a generic term including all specialized or restricted types of facies. *See* STRATIGRAPHY.

The term facies is widely used in connection with sedimentary rock bodies, but is not restricted to these. In sedimentary rocks, the changes in facies reflect changes in environments of deposition. Thus, fresh-water environments in rivers change to brackish-water environments in estuaries and to marine conditions in the ocean. In general, facies are not defined for sedimentary rocks by features produced during weathering, metamorphism, or structural disturbance. In metamorphic rocks, specifically, however, facies may be identified by the presence of minerals that denote degrees of metamorphic change. *See* PETROLOGY.

Lithofacies. In geological studies, the facies most commonly designated are lithofacies, defined as a lateral subdividing of a stratigraphic rock body differentiated from other adjacent subdivisions by its lithologic characteristics. Lithofacies may be separated laterally by vertical cutoff planes that divide the rock unit into sharply differentiated classes or by interfingering surfaces that produce intergrading types. This kind of facies is a mappable subdivision of the rocks.

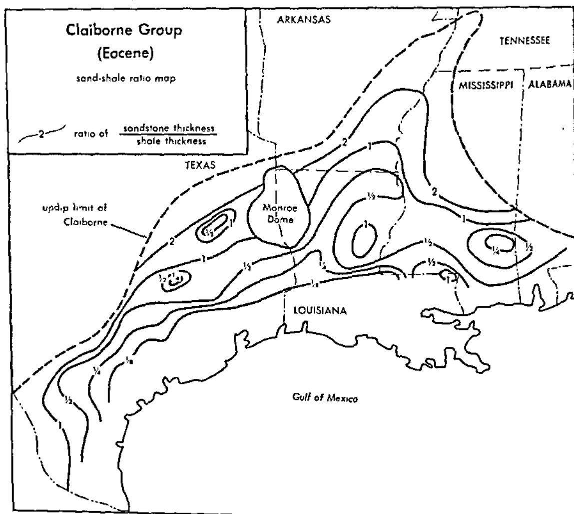
Mappable lithofacies can be differentiated on any selected lithologic basis. One basis is gross lithologic composition, such as the percentages of sandstone, shale, and limestone, or the ratio of any of these rock types to the others. A second basis may be specific lithologic attributes such as colors of shale or varieties of limestone types. The differentiation may also be based on combinations of rock attributes that designate conditions of sediment deposition, such as terrestrial, marine, or geosynclinal facies. *See* SEDIMENTARY ROCKS; SEDIMENTATION (GEOLOGY).

Lithofacies maps show the areal distribution of the several differentiated aspects. Such maps find application in oil exploration, where they are used in conjunction with structure-contour maps and maps of the total thickness of the stratigraphic unit (isopach maps). The isopach and structure maps convey information on the over-all geometry of the stratigraphic rock body. The isopach map shows its areal extent and thickness variations; the structure map shows the structural attitude of its upper surface. The lithofacies map gives information on the changing composition of the rock body throughout its geographic occurrence. This changing composition may indicate areas favorable for oil and gas exploration, inasmuch as certain combinations of sedimentary rocks tend to occur in known areas of oil production. In combination with structurally high areas, such as domes and anticlines, changes in facies may be important factors in localization of oil or gas. Stratigraphic oil fields, in which the controlling factors are the facies attributes themselves, may sometimes be located on a regional basis by means of lithofacies maps. Similarly, other kinds of mineral deposits associated with sedimentary rock bodies, such as uranium, lead, and zinc ores, and some others, commonly bear some relation to the facies patterns of the sedimentary rocks. *See* PETROLEUM GEOLOGY; PROSPECTING, PETROLEUM.

Lithofacies studies also shed light on the historical geology of sedimentary rocks because they provide information which is useful in determining the conditions under which the sediments were laid down, such as shallow or deep marine waters, deltaic conditions, tidal lagoons, or other conditions of deposition. They also provide some indication of the broad features of relative subsidence in the earth's crust during accumulation of the deposits.

Biofacies. Lithofacies studies contribute much to the knowledge of past environments; however, certain aspects of those environments can be unraveled only by studies of the fossil organisms contained in the sedimentary rock body. Lateral subdivision of a stratigraphic unit differentiated from other adjacent subdivisions by its biological characteristics constitutes a *biofacies*. Biofacies maps may be based on proportions or ratios among the fossil organisms, much as lithofacies maps are based on lithologic attributes. They are particularly useful in reconstructing the environment of deposition, inasmuch as the organisms commonly reflect conditions of water depth, salinity, temperature, or other conditions. *See* PALEOECOLOGY.

Facies maps. The importance of facies changes in rocks, in the reconstruction of geologic history, and in the exploration of ore and mineral depos-



Lithofacies map based on ratio of two rock components. (*After R. G. Strand*)

its, has given rise to a wide variety of methods for preparing facies maps. Most maps are prepared somewhat in the manner of a contour map by drawing lines of equal magnitude through a field of numbers representing the observed values of the measured rock attributes. The basic lithofacies data are obtained by making thickness measurements of the several kinds of rocks penetrated by wells in the mapped area or exposed in outcrops along valley walls. In the simplest kind of facies map (sometimes called an isolith map), the total thickness of a selected rock type in the stratigraphic unit, such as a sandstone, is plotted at each point of observation, and contours of net sand thickness are drawn through the numbers. Such a map shows the absolute quantity of sandstone in the stratigraphic unit, and may prove useful in study of the distribution of possible oil reservoirs in an area of interest. Maps of relative rock content, such as the percentage of sandstone in the unit or of the ratio of sandstone to shale, give other kinds of information about the same rock body. Commonly several lithofacies maps are prepared for the unit under study, to provide information on more than one lithologic attribute.

The accompanying illustration is a lithofacies map of the Claiborne Group (Eocene) of the Gulf Coast region, showing the ratio of sandstone to shale in that body of rock. The ratios are contoured on a geometric basis and show a rapid decrease to the south. In the north there is more than 2 ft of sandstone per foot of shale, whereas near the southern limit of available data the amount of sandstone per foot of shale is less than ⅛ ft. The map reflects the commonly observed decrease in sand as mixtures of sand and mud are redistributed by waves and currents along ancient shorelines and into deeper water offshore. Irregularities in the subparallel facies lines indicate the influence of local factors, usually variations in water depth or structural growth in the depositional area.

Another application of ratio techniques in facies mapping is to contrast the detrital portions of a sedimentary body (the sand and mud carried into the depositional site by streams or other agencies) with nondetrital sediments, such as limestone, that may be formed directly within the environment by chemical or biological processes. This contrast may be expressed as the clastic ratio, the ratio of the combined thicknesses of sandstone and shale in

the unit to the thickness of limestone. In this case the ratio indicates the number of feet of sand and mud deposited per foot of limestone. It is a measure of the relative contribution by materials carried into the environment compared to sediments formed locally. Such combination ratios provide a means for showing the mutual relations among several lithologic components by a single set of contour lines. For an illustration of lithofacies mapping techniques utilizing the clastic ratio *see* SILURIAN. [W.C.K.]

Bibliography: W. C. Krumbein and L. L. Sloss, *Stratigraphy and Sedimentation*, 1951; J. M. Weller, Stratigraphic facies differentiation and nomenclature, *Bull. Am. Assoc. Petrol. Geologists*, 42:609-639, 1958; L. L. Sloss, E. C. Dapples, and W. C. Krumbein, *Lithofacies Maps*, 1960.

Facsimile

The transmission of fixed images by wire or radio The images or subject copy are usually in the form of photographs, written messages, maps, or drawings.

The largest application of facsimile is in message telegraphy, with more than 40,000 instruments in service in the United States. Approximately 1000 facsimile sets are used to record weather maps Newspapers and television stations employ facsimile for the recording of newsphotos. New uses that are expected to prove important include the transmission of x-ray photos and the high-speed transmission of newspaper and magazine proof pages from which printing plates are made.

Transmission system. The subject copy is divided into dots, or elemental areas, in the scanning and transmitting process. Each elemental area (1/100 by 1/100 in. in a typical scanning system) is transmitted as an electrical pulse. If an amplitude-modulated transmitting system is employed, the electrical pulse usually represents the blackness of each elemental area.

In an amplitude-modulated system, the impulses are transmitted at a constant rate, usually 1800, 2000 or 2400 pulses/sec when transmission is over a voice-frequency circuit. The pulse rate is referred to as the carrier frequency.

Some circuits cannot retain the true amplitude relationship of the pulses in transmission. For example, high-frequency radio circuits are subject to considerable fading. In such cases amplitude modulation is unsatisfactory and frequency-shift modulation is used so that the received copy will not be degraded by fading or level changes.

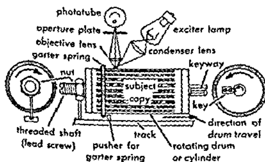


Fig. 1. Rotary facsimile scanner.

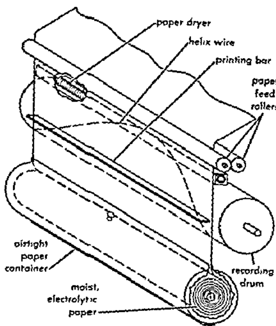


Fig 2. Electrolytic facsimile recorder.

The received signals are recorded on a record sheet by any one of many processes. In a direct-recording facsimile system the images are visible while being recorded. In photofacsimile systems employing photographic film or paper, photographic processing is necessary before the image may be viewed.

The transmitted elemental areas are placed in the correct positions on the record sheet by properly synchronizing the scanning and recording mechanisms. The two mechanisms are driven at the same speed by synchronous motors. The receiving motor must be supplied by electrical power that is synchronous with the power supplying the transmitter motor. When the transmitting and receiving stations are supplied by the same power system, the synchronism at the two separate locations is usually sufficiently close and stable. If not, or if different power systems are involved, local power is generated at each station. Exceptionally precise control of the frequency is necessary, usually accomplished by a tuning-fork oscillator.

Scanning. A phototube views the elemental areas through an optical system generally consisting of an objective lens that forms an image on an aperture plate (Fig. 1). The aperture in this plate permits only the light from an elemental area to pass to the phototube. In order that the phototube may scan the entire area of the subject copy, the copy may be placed on a rotating cylinder and the drum slowly moved the length of its shaft so as to scan a line about 1/100 in. wide each turn. Several mechanical equivalents are employed. If the co

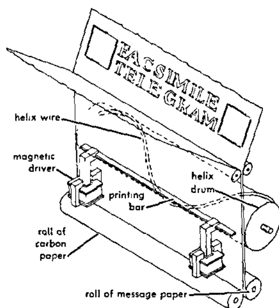


Fig 3. Pressure-sensitive mechanical recorder.

to be scanned in the flat (flat-bed scanner), the beam to the phototube is optically swept across the width of the copy as the copy is slowly advanced. In one variation of this, an illuminating beam sweeps across the copy and the phototube picks up the reflected light.

Modulation. The carrier frequency may be developed by interrupting the beam of light striking the phototube. The beam intensity varies with the light picked up from the elemental areas so that an amplitude-modulated signal is developed directly by the phototube If the beam is not modulated, the dc picture signals from the phototube may modulate a carrier signal in a conventional modulator, preferably one that suppresses the modulating frequency. A third system uses a phototube as part of a balanced bridge modulator.

If frequency is to represent the blackness of the elemental areas, the amplified AM picture signal is converted back to a dc picture signal. This demodulated signal may control the frequency of an audio oscillator, 1500 cps representing white and 2300 cps representing black. These signals may be transmitted over an amplitude-modulated radio transmitter. This type of modulation is referred to as audio-frequency shift. The dc picture signal may serve to shift the radio carrier frequency over a range of 800 cps. At the receiving station, a heterodyne develops the 1500–2300-cps audio-frequency signal, which is passed through a limiter and discriminator ahead of the facsimile recorder.

Direct recording. Direct recording is usually accomplished by one of three techniques. A current may be passed through a paper dampened with an electrolyte (Fig. 2). A white coating covering a conducting carbon layer on paper may be burned off electrically. The elemental areas may be printed mechanically (Fig. 3). A drum, on which there is a wire helix, rotates in synchronism with the cylinder of the transmitter. A printing bar strikes the carbon paper whenever a black pulse is received, thereby printing on the message paper at the location determined by the instantaneous position of the helix wire.

Synchronization. If the synchronous motors are driven by local oscillators, the frequency difference between the transmitting and receiving oscillators should not exceed .001% when the system is transmitting letter-size message copy (scanned at 100 lines in); 0003% when transmitting 12- by 18-in. weather maps (scanned at 100 lines/in.); and .00004% when transmitting 16- by 22-in. newspaper copy scanned at the rate of 400 lines/in. (This scanning rate is required for high definition.) Tuning-fork oscillators are most commonly employed because of their long-term stability as compared with crystal oscillators. *See* TELEGRAPHY; TELEPHOTOGRAPHY. [A.C.C.]

Fagales

An order of the plant subclass Dicotyledoneae including 2 families, the birch family (Betulaceae), with 6 genera and 105 species, and the beech family (Fagaceae), with 6 genera and 600 species. They are shrubs or trees occurring mostly in the north temperate and subtropical zones, but also having representatives in the tropical and south temperate regions The birch family includes hornbeams (*Carpinus*) and the hophornbeams (*Ostrya*) of Europe, America, and Japan; the Tibetan, Chinese, European, and American hazels (*Corylus*); the tall birches (*Betula*) of Europe and eastern North America, the dwarf birches of the Arctic, and the alders (*Alnus*). The beech family, including the beech, oak, and chestnut, produces lumber of great economic value. *See* BEECH; BIRCH; CHESTNUT; HOPHORNBEAM; HORNBEAM; OAK; *see also* DICOTYLEDONEAE; EMBRYOPHYTA; PLANT KINGDOM; TREE. [P.D S.]

Falconiformes

The order containing the diurnal birds of prey. There are 7 families, 5 living and 2 extinct. One of the fossil species, *Teratornis incredibilis* from the late Pleistocene of Nevada, is calculated to have had a wingspread of up to 17 ft, the largest flying bird thus far known. Among living species the largest is the Andean condor (*Vultur gryphus*), which attains a 12-ft wingspread, whereas the Philippine falconet (*Microhierax erythrogonys*) is scarcely larger than a sparrow. Most members of this order are rapacious, such as the hawks, eagles, kites, falcons, and similar birds which are adapted for capturing living prey ranging in size from insects to small deer. Some are primarily carrion-feeders, including the New World vultures (Cathartidae), Old World vultures (Aegypiinae), and caracaras (Falconidae, subfamily Daptriinae). Both of the extinct families were also vulturelike Two somewhat specialized species are each given family rank, the widely distributed fish-eating osprey (Pandionidae) and the long-legged, primarily

terrestrial secretary bird of Africa (Sagittariidae). The diurnal birds of prey are no longer considered to be related to the owls. Similarities in the two predacious groups are the result of convergence. *See* AVES. [K.C.P.]

Fall line

The zone or boundary between resistant rocks of older land and weaker strata of plains. Here the easily eroded strata of the plains lap onto the more resistant rocks of the old land (Fig. 1). Streams that cross this boundary have steeper gradients along the width of the zone and make sudden descents over falls and rapids. In geologic and geographic literature, a line connecting falls of this type on several main rivers is referred to as a fall line.

In the eastern United States, the boundary between the Coastal Plain and the Piedmont geographic section to its west (Fig. 2) has long been known as the Fall Line, and serves as the origin for the general term. Where such rivers as the Delaware, Potomac, James, Roanoke, and Savannah cross the boundary zone there are falls or rapids that limit upstream navigation and provide a source of water power. Some geologists believe that this

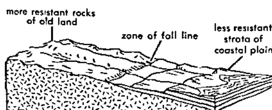


Fig 1. Diagram showing fall line zone.

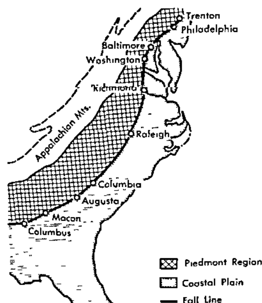


Fig. 2. The fall line marks the boundary between the Coastal Plain and Piedmont region in eastern United States.

fall line represents the surface trace of an ancient peneplain (old erosion surface) of pre-Cretaceous age (the Fall Line Peneplain) now buried beneath younger strata of the coastal plain. A similar, now-buried fall line, with falls that once attained heights of as much as 80 ft. lies beneath alluvial fill along the east scarp of the Ozark Plateau. *See* COASTAL PLAIN; FLUVIAL EROSION CYCLE. [S.J.]

Fallopian tube

A part of the female reproductive system present in the higher vertebrates. The Fallopian tube passes the ovum from the ovaries to the cavity of the uterus. Each tube is about 5 in. in length; one lies on either side of the uterus and is attached at the upper portion. Each curves outward to end in a structure with many fingerlike projections, the fimbriae. The ovaries lie below and inside the tubal curve. When ovulation occurs, about the middle of each menstrual cycle (in human beings), the ovum is picked up by the fimbriae and drawn into the oviduct. The lining of the tube is epithelium which has tiny hairlike processes, the cilia, on its surface. These cilia, and probably muscular action, move the ovum along the tube. The ovum is viable for about 1–3 days only. If fertilization does occur, the ovum then normally moves along to the uterine cavity where it becomes implanted on the wall. If fertilization does not occur, the ovum degenerates in the uterus. Occasionally, a fertilized ovum fails to move into the uterus, or may be freed into the abdominal cavity, so that an ectopic pregnancy results if the ovum finds a site for successful implantation. *See* PREGNANCY, DISORDERS OF.

Comparable structures derived from the primitive oviducts exist in many lower forms, but are most characteristically present in placental mammals. *See* REPRODUCTIVE SYSTEM. [E.G ST.]

Fan

A fan moves gases by producing a low compression ratio as in ventilation and pneumatic conveying of materials. The increase in density of the gas in passing through a fan is negligible (no more than 7%); the pressure increase or head is usually measured in inches of water.

Blowers are fans that operate where the resistance to gas flow is predominantly downstream of the fan. Exhausters are fans that operate where the flow resistance is predominantly upstream of the fan.

Fans are further classified as centrifugal or axial (Fig. 1). The housing provides an inlet and an outlet and confines the flow to the region swept out by the rotating impeller. The impeller imparts velocity to the gas, and this velocity changes to a pressure differential under the influence of the housing and ducts connected to inlet and outlet.

Performance. In selecting a fan for a particular application, requirements of primary interest are the quantity Q of gas to be delivered by the fan ar the head H against which the quantity Q of gas fan capacity must flow. These operating con

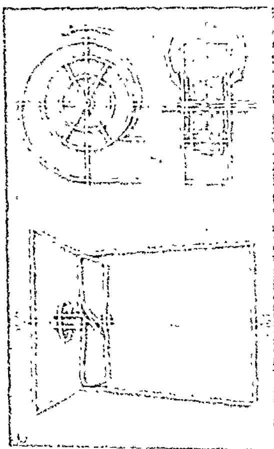


Fig. 1 Two principal types of fan. (a) Centrifugal. (b) Axial or propeller.

establish the fan dimensions of diameter D and rotational speed N. Performance of a fan of diameter D is rigorously described by its characteristic curves (Fig. 2). Fans of different types and sizes are conveniently compared by converting their characteristics to dimensionless form (Fig. 3).

The load placed on the fan must correspond to a condition on its operating characteristic. Thus if, for example, the system in which the fan operates presents less resistance than is overcome by the head developed by the fan at the required capacity, additional resistance must be introduced by a damper, or the fan speed must be changed, or the excess capacity must be diverted elsewhere.

Performance of a given fan, operating at a given point on its efficiency curve, varies with speed in accordance with the following rules:

Capacity Q proportional to speed N
Head H proportional to speed squared N^2
Horsepower P proportional to speed cubed N^3

For a series of similar fans, operating at a given point on their efficiency curves and at the same speed, performance varies in accordance with these rules:

Capacity Q proportional to diameter cubed D^3
Head H proportional to diameter squared D^2
Horsepower P proportional to diameter to fifth power D^5

Dimensions and performance of a selected group of fans

Fan dimensions and performance	Centrifugal fans: Backwardly curved blades	Centrifugal fans: Steel plate	Centrifugal fans: Forwardly curved blades	Axial fan: Pressure blower
Performance				
Specific speed, N_s	0 25	0.11	0 21	0 75
Peak efficiency, %	75	65	65	72
Tip speed, fpm × 10^3	18	11	3 5	16
C_H at shut-off	3 8	6.3	11.2	3 2
C_Q at peak eff	0 42	0.18	1.2	0 4
C_P at peak eff	2 1	1.7	2 1	0 55
Proportions as functions of wheel diameter, D				
Inlet area (XD^2)	0 7	0.4	0 8	0 8
Outlet area (XD^2)	0 6	0 3	0.7	
Axial blade length (XD)	0 3	0.4	0 6	
Number of blades	18	8	64	6

These six relations can be combined to express the performance of a family of similar fans in terms of dimensionless coefficients as follows:

$$\text{Capacity coefficient } C_Q = Q/ND^3$$
$$\text{Head coefficient } C_H = gH/N^2D^2$$
$$\text{Power coefficient } C_P = gP/wN^3D^5$$

where Q = capacity, ft³/sec
H = head, ft of fluid
P = shaft horsepower, ft-lb/sec
g = gravitational acceleration, ft/sec²
N = revolutions/sec
w = weight density, lb/ft³
D = wheel diameter, ft

Values of these coefficients for illustrative type fans are listed in the table.

Comparison of different types of fans is further facilitated by the elimination of the diameter term. Equations for capacity coefficient and head coefficient are solved simultaneously to give specific speed N_S as

$$N_S = NQ^{1/2}/(gH)^{3/4}$$

Specific speed is an inherent performance criterion; it is usually employed as a dimensionally impure coefficient. In the selection of a fan for a

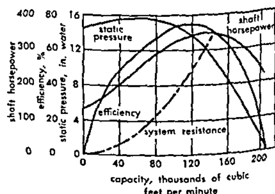


Fig. 2. Static pressure characteristics of backwardly curved blade centrifugal fan at constant speed.

given application, a fan is chosen that has the desired value of N_S in its region of peak efficiency. Low specific speed corresponds to a fan of low rotative speed and large impeller diameter.

Alternatively speed can be eliminated, giving specific diameter D_S as

$$D_S = D/Q^{1/2}$$

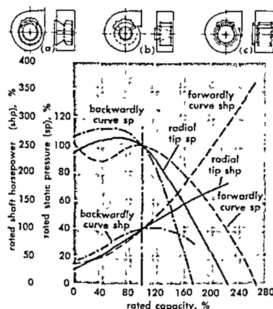


Fig. 3. Normalized characteristics compare performance of three forms of centrifugal fans.

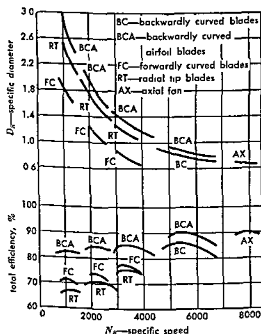


Fig 4 Comparison of fan types on basis of specific [illegible]

Characteristics of fans are compared on the basis of these ratios (Fig. 4). An application determines the capacity and head that a fan must provide. A fan is selected of a size that has the required capacity and head at the rotational speed of the available drive.

Types. The shape of a fan characteristic as a function of capacity depends on the fan type. For example, backwardly curved blade centrifugal fans have a steeply falling head characteristic and a self-limiting horsepower characteristic (Fig. 2). Fans with forwardly curved blades have rising head and horsepower characteristics. The axial fan has a rising horsepower characteristic so that at shutoff the fan may require more power than at its useful flow rate. *See* DUCTED FLOW; PUMPING MACHINERY [T.B.]

Bibliography: T. Baumeister, *Fans*, 1935; W. H. Carrier, *Fan Engineering*, 4th ed., 1938; A. H. Church, *Centrifugal Pumps and Blowers*, 1944; G. F Wislicenus, *Fluid Mechanics of Turbomachinery*, 1947.

Farad

The unit of capacitance in the meter-kilogram-second (mks) system of units. A capacitor has a capacitance of 1 farad if there is a potential difference of 1 volt between its plates when the charge on one of the plates is 1 coulomb (there being an equal and opposite charge on the other plate). Thus, 1 farad = 1 coulomb/volt. *See* CAPACITANCE; ELECTRICAL UNITS. [R.P.WI.]

Faraday effect

Rotation of the plane of polarization of a beam of linearly polarized light when the light passes through matter in the direction of the lines of force of an applied magnetic field. Discovered by M. Faraday in 1846, the effect is often called magnetic rotation. The magnitude α of the rotation depends on the strength of the magnetic field H, the *nature of the transmitting substance, the frequency* ν of the light, the temperature, and other parameters. In general

$$\alpha = VxH$$

where x is the length of the light path in the magnetized substance and V the so-called Verdet constant The constant V is a property of the transmitting substance, its temperature, and the frequency of the light.

The Faraday effect is particularly simple in substances having sharp absorption lines, that is, in gases and in certain crystals, particularly at low temperatures Here the effect can be fully explained from the fundamental properties of the atoms and molecules involved. In other substances the situation may be more complex, but the same principles furnish the explanation.

Rotation of the plane of polarization occurs *when there is a difference between the indices of* refraction n^+ for right-handed polarized light and n^- for left-handed polarized light (*see* POLARIZED LIGHT). Most substances do not show such a dif-

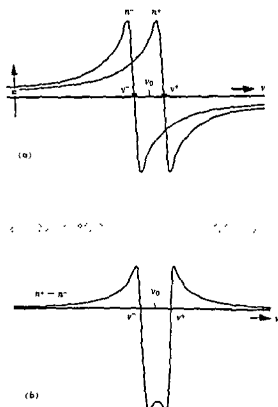


Fig 1. Curves used in explaining the Faraday effect, where the quantity ν is the frequency of the light (a) Index of refraction for left-handed circularly polarized light (n^-) and right-handed light (n^+) in the vicinity of an absorption line split into a doublet (ν^-,ν^+) in a magnetic field. (b) The difference between the two curves, $n^+ - n^-$ The magnetic rotation is proportional to this difference.

ference without a magnetic field, except *optically* active substances such as crystalline quartz or a sugar solution. It should be noted that the index of refraction in the vicinity of an absorption line changes with the frequency, as shown by Fig. 1*a* *See* ABSORPTION (ELECTROMAGNETIC RADIATION).

When the light travels parallel to the lines of force in a magnetic field, an absorption line splits up into two components which are circularly polarized in opposite directions; this is the normal Zeeman effect (*see* ZEEMAN EFFECT). This means that for one line, only right-handed circularly polarized light is absorbed and for the other one, only left-handed light. The indices of refraction n^- and n^+ bear to their respective absorption frequencies the same relation as indicated in Fig. 1*a*; that is, they are identical in shape but displaced by the frequency difference between the two Zeeman components. It is evident that $n^+ - n^-$ is different from zero (Fig. 1*b*), and the magnetic rotation is proportional to this *difference*. The magnitude of the rotation is largest in the immediate vicinity of the absorption line and falls off rapidly as the frequency of the light increases or decreases.

The Faraday effect may be complicated by the fact that a particular absorption line splits into more than two components or that there are several original absorption lines in a particular region of the spectrum.

The case represented in Fig. 1 is independent of the temperature, and the rotation is symmetric on both sides of an absorption line. This case is called, not quite correctly, the diamagnetic Faraday effect. It occurs when the substance is diamagnetic, which means the splitting of the absorption line is due to the splitting of the upper level only (Fig. 2*a*), and the lower level of the line is not split. The same situation prevails in general when the intensity of the two Zeeman components is equal, which is the case for all substances except paramagnetic salts at very low temperatures (Fig. 2*b*).

In the latter case at high temperatures there are an equal number of ions in the +1 and −1 states and the two Zeeman components have equal intensities, as in the previously discussed situation. When the temperature is lowered, however, the ions concentrate more and more in the lower level (−1) and therefore absorption by ions in the upper level (+1) disappears. At very low temperatures only the high-frequency component in the Zeeman pattern is left. In this case the n^- refraction coefficient is not affected by the presence of the absorption line and is a constant. The difference $n^+ - n^-$ will therefore have the shape of the

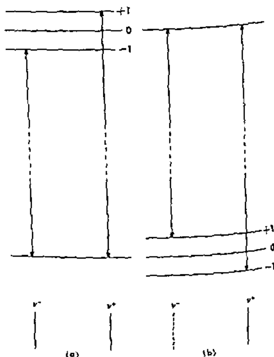


Fig. 2. Two cases of the Faraday effect. (a) The so-called diamagnetic case, which is temperature-independent and (b) the paramagnetic case, which is temperature-dependent. At very low temperatures the ν^- component in (b) is absent.

n' curve in Fig. 1a. Here the rotation is not symmetric with respect to the absorption line. If it is right-handed on one side of the line, it will be left-handed on the other. As the temperature is raised, the other line comes in with increasing strength until the two are nearly equal. In the transition region the Faraday effect depends strongly on the temperature. This case has been called the paramagnetic Faraday effect.

For a discussion of other phenomena related to the Faraday effect see MAGNETOOPTICS. [C.H.D.]

Faraday's law of induction

A statement relating an induced electromotive force (emf) to the change in magnetic flux that produces it. For any flux change that takes place in a circuit, Faraday's law states that the magnitude of the emf $\mathcal{E}$ induced in the circuit is proportional to the rate of change of flux.

$$\mathcal{E} \propto -\frac{d\Phi}{dt}$$

The time rate of change of flux in this equation may refer to any kind of flux change that takes place. If the change is motion of a conductor through a field, $d\Phi/dt$ refers to the rate of cutting flux. If the change is an increase or decrease in flux linking a coil, $d\Phi/dt$ refers to the rate of such change. It may refer to a motion or to a change that involves no motion.

Faraday's law of induction may be expressed in terms of the flux density over the area of a coil. The flux Φ linking the coil is

$$\Phi = \int B \cos \alpha \, dA$$

where α is the angle between the normal to the plane of the coil and the magnetic induction B. The integral is taken over the area A enclosed by the coil. Then, for a coil of N turns,

$$\mathcal{E} = -N\frac{d\Phi}{dt} = -N \int \frac{d(B \cos \alpha)}{dt} A$$

See INDUCTION, ELECTROMAGNETIC; MAGNETIC FLUX. [K.V.M.]

Farinales

An order of the plant subclass Monocotyledoneae characterized by the mealy endosperm of the seeds. The validity of this group has been questioned, but as yet there is no general acceptance of any proposed redistribution of the five families involved. These are mainly tropical and subtropical plants, with some representatives in temperate regions. Some well-known members are yellow-eyed grass, pipewort, bog-buttons, day flower, wandering Jew, pineapple, and Spanish moss (a misnomer).

The pineapple family (Bromeliaceae), with 65 genera and over 1500 species, is much the largest family including the most conspicuous epiphytes of the tropical rainforest. See PINEAPPLE; see also EMBRYOPHYTA; MONOCOTYLEDONEAE; PLANT KINGDOM. [P.D.S.]

Fascioliasis

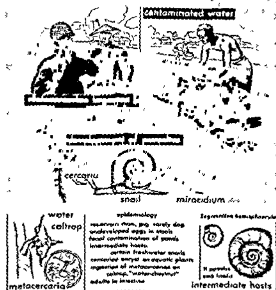
The accidental infection of men with *Fasciola hepatica*, the cattle liver fluke. The leaflike fluke inhabits the bile ducts and is a world-wide agricultural scourge. Infected animals pass the eggs in feces. The eggs embryonate in water and the miracidium hatches in 2–3 weeks and then penetrates into amphibious lymnaeid snails. After sporocyst and rediae stages, cercariae finally arise, emerge, and encyst on vegetation, in water, or on submerged obstacles. The parasite thus prevails in swampy pastureland.

Man is exposed by accidentally carrying contaminated objects to his mouth. Incidence is low, only several hundred cases being known. Symptoms are directly referable to biliary dysfunction and may be confused with various liver diseases. Emetine hydrochloride is considered specific for treatment. See PARASITOLOGY, MEDICAL; TREMATODA. [J.F.MA.]

Fasciolopsiasis

The presence of the trematode *Fasciolopsis buski* in man's small intestine. The large and fleshy parasite is native to Eastern Asia and the Southwest Pacific. Hogs act as natural reservoirs. Man becomes parasitized by using his teeth to peel various aquatic raw vegetables on which the parasite encysts after its development in a molluscan host





Epidemiology of Fasciolopsiasis. (From T. T. Mackie, G. W. Hunter, and C. B. Worth, A Manual of Tropical Medicine, 2d ed., Saunders, 1954)

The fluke lives attached to the intestinal mucosa by its two powerful suckers, producing inflammation and ulceration. A light infection is characterized by toxic diarrhea. A heavy one produces toxic symptoms leading to edema and heart failure.

The disease is diagnosed by the recovery of eggs in the feces and treatment is 0.4–1.0 g of hexylresorcinol by mouth. See PARASITOLOGY, MEDICAL; TREMATODA. [J.F.MA.]

Fat and oil, edible

One of the three major classes of food products. Along with carbohydrates and proteins, fats and oils supply the energy requirements of man and animals. They consist principally of glyceride esters of fatty acids and are characteristically soluble in organic solvents such as petroleum hydrocarbons, ether, chloroform, and carbon tetrachloride. Fats are usually defined as solid or plastic at ordinary temperatures, but the term may be used for all such compositions regardless of melting point. Oils are *liquid*. The two major classes are

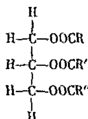
processed to obtain edible fat. Butter is a special type of animal fat product from milk. The *vegetable oils* are pressed or extracted from a variety of plant seed. Of primary importance as *sources of* edible oil on a world basis are soybeans, cottonseed, peanuts, corn germ, olives, coconut, rapeseed, sesame, sunflower seed, cocoa beans, and various oil palms. Major domestic production is limited to the first four of these, but dried coconut meat, that is, copra, is imported and processed domestically and imported cocoa beans are processed

mated world production of food fats and oils exceeds 36,000,000,000 lb. Fats and oils *in the diet* serve to increase palatability and enhance the flavor of foods, provide a lubricating action *and improve* the texture of baked goods, increase the satisfaction of eating, and delay the onset of hunger.

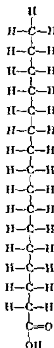
Nutritive value. The nutritive value *of fats* and oils is characterized by high energy content. The

elements, the *percentage* of ingested fat absorbed, range from 94 to 98%. Fats *that melt* above 50°C, such as mutton fat, are less *completely* digested. Fats and oils serve as carriers *for the fat*-soluble vitamins A and D, aiding their *absorption* from the intestinal tract, and are the chief *source of vitamin E*. They also have a sparing action on some vitamins of the B complex. *They* are the source of the polyunsaturated or *essential* fatty acids, required for structural development *of* tissues and prevention of fat-deficiency disease which in man manifests itself as eczema. See ECZEMA; NUTRITION; VITAMIN A; VITAMIN D; VITAMIN E.

Chemical constitution. Edible fats and oils are basically esters of glycerol, $C_3H_5(OH)_3$, and various fatty acids having an even number of carbon atoms arranged in a long straight chain. They may be represented by the formula



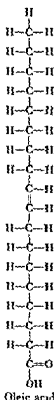
where R, R′, and R″ represent the carbon chains of the fatty acids. The fatty acids may be saturated, that is each carbon atom in the chain is linked by single bonds to other carbon atoms or to hydrogen atoms as in palmitic acid



Palmitic acid

or may be unsaturated, having one or more carbon atoms in the chain joined by two bonds as in oleic acid.

Hydrogen, halogens, and other chemical reagents can be added to the double bonds of unsaturated acids to form saturated acids or derivatives. The most important saturated acids in edible fats are palmitic with 16 carbon atoms, stearic with 18 carbon atoms, lauric with 12 carbon atoms, and butyric with 4 carbon atoms. Oleic acid, having 18 carbon atoms and a double bond in the 9 carbon atom position, is the most abundant unsaturated acid. Acids having more than one double bond include linoleic, having 18 carbon atoms and 2 double



Oleic acid

bonds in the 9 and 12 positions; linolenic acid, 18 carbon atoms with three double bonds in the 9, 12, and 15 positions; and arachidonic acid, having 20 carbon atoms and 4 double bonds in positions 5, 8, 11, and 14. The amounts of different fatty acids constituting various fats are shown in the table. *See* LIPID.

The saturated and unsaturated fatty acids of vegetable oils may be evenly or randomly distributed among the glyceride molecules, resulting in a complex mixture. In animal fats considerable proportions of fully saturated glycerides occur. Only in a few oils such as cocoa butter are there a very limited number of glycerides of specific configuration. Cocoa butter consists almost exclusively of two glycerides, 2-oleopalmitostearin and 2-oleodistearin.

Typical fatty acid composition of animal fats and vegetable oils[a]

	Saturated			Unsaturated		
	Palmitic	Stearic	Other[b]	Oleic	Linoleic	Other
Fats						
Lard	29.8	12.7	1.0	47.8	3.1	5.6[c]
Chicken	25.6	7.0	0.3	39.4	21.8	5.9[c]
Butterfat	25.2	9.2	25.6	29.5	3.6	7.2[c]
Beef fat	29.2	21.0	1.4	41.1	1.8	3.5[c]
Oils						
Corn	8.1	2.5	0.1	30.1	56.3	2.9
Peanut	6.3	4.9	5.9	61.1	21.8	
Cottonseed	23.4	1.1	2.7	22.9	47.8	2.1
Soybean	9.8	2.4	1.2	28.9	50.7	7.0[d]
Olive	10.0	3.3	0.6	77.5	8.6	
Coconut	10.5	2.3	78.4	7.5	Tr	1.3
Cocoa butter	24.0	35.0		39.0	2.0	

[a] As weight percentages of component fatty acids.
[b] Butterfat and coconut oil contain fatty acids of less than 12 carbon atoms. Peanut oil contains several per cent of acids of 20 carbon atoms and above.
[c] Mainly hexadecenoic acid; 0.2–0.4% arachidonic acid.
[d] Mostly linolenic acid.

Nonglyceride components which are not saponified by alkali constitute from about 0.5 to 2% of most fats. They include sterols, hydrocarbons such as squalene, carotenoids, and fat-soluble vitamins. The tocopherols (one of which is vitamin E) function as antioxidants and delay the development of rancidity in the fat.

Spoilage factors. Oxidation and hydrolysis are spoilage factors in the production and storage of edible fats and oils. Incipient oxidation may produce flavors characterized as grassy, buttery, beany, or fishy. Rancidity is an advanced state of oxidative deterioration. Oxygen from the air first reacts with the unsaturated fatty acids at or adjacent to the double bonds to form hydroperoxides which then decompose to yield aldehydes having the pungent odor and flavor of rancid fats. Oxidation is catalyzed by light and metals such as copper or iron, and is accelerated by heat. Preventive measures include packaging in brown glass or metal containers; use of nitrogen, an inert gas, in processing and packaging; addition of citric acid during processing to inactivate trace metals; and addition of antioxidants such as NDGA (nordihydroguiaretic acid), BHA (butylated hydroxyanisole), BHT (butylated hydroxytoluene), or propyl gallate to fats deficient in naturally occurring antioxidants. The more-saturated fats and hydrogenated fats are less subject to oxidative deterioration. In processing or use, fats should never be in contact with equipment or utensils made of copper or copper-containing alloys.

Hydrolytic spoilage results only in fats in contact with moisture. Butter, margarine, and many processed foods contain enough moisture that hydrolysis may be encountered. With most fats hydrolysis does not noticeably affect the flavor, but butter becomes strong and coconut oil develops a soapy flavor. This type of spoilage may be catalyzed by the enzymes present in other components of a processed food product or enzymes liberated by microorganisms.

Testing procedures. Tests used in establishing quality and grade of crude vegetable oils include free fatty acid content, determined by titration of an alcohol solution of the oil with standard alkali, and refining and bleaching tests. Replicated refining tests on 500-g oil samples are carried out using specified amounts of sodium hydroxide and the refining loss determined by weighing the soapstock and the decanted refined oil. The color of the refined oil, or oil bleached by heating with a standard bleaching clay, is determined photometrically.

Various characteristics of processed oil and fat products are used in specifications and control

Smoke points are determined on cooking oils by heating in an open cup to the smoking temperature of about 440°F or above. Salad oils must pass at least a 5½-hour cold test during which the oil must not cloud or crystallize when packed in ice.

Unsaturation is measured as iodine value, calculated as the per cent iodine absorbed by the fat. As determined by the Wijs method, a weighed sample of fat is dissolved in carbon tetrachloride and 100–150% excess of iodine monochloride in glacial acetic acid is added. After ½ hour, potassium iodide is added to convert the unused reagent to iodine which is titrated with sodium thiosulfate. Iodine value is related to hardness of a fat and its stability to oxidation, the lower the iodine value the harder and more stable the fat.

Stability to oxidation is measured by accelerated tests at elevated temperatures. In the Schaal test the fat is heated in an oven at 140°F until rancidity develops. Other tests involve aerating the heated fat under standardized conditions and periodically measuring peroxide formation by iodometric titration or measurement of oxygen absorption of fat held at a constant elevated temperature. With specified fat products these accelerated tests may be correlated with shelf life.

Plastic properties of shortening, margarine oil, and coating fats are very important and a number of methods are useful for defining consistency or plastic range. Micropenetration is a method which measures in tenths of a millimeter the penetration of a steel needle dropped on the fat conditioned overnight by chilling in a refrigerator and then tempering at the tempe[illegible] M[illegible] a[illegible] cropenetration values at 25°C for shortening, margarine fat, and lard are 10.0, 11.7, and 13 7 mm, respectively.

Plastic properties are dependent upon the content of liquid and solid glycerides. The percentage of solid glycerides at any given temperature may be estimated by measuring the expansion of the fat confined in a dilatometer as the temperature is raised and the volume changes with change of state from solid to liquid.

Other characteristics frequently determined include moisture and volatiles, saponification value, unsaponifiable matter, density, refractive index, melting point, and titer. For shortenings containing emulsifiers, monoglyceride content is determined iodometrically following oxidation with periodic acid.

Production methods. Processing of oilseeds is carried out by pressing, extraction with solvents, or a combination of the two. Preparation of the seed requires cleaning, hulling, and separation of the hulls on shaking screens and by aspiration. The meats are ground or cracked and flaked to a thickness of 0.005–0.010 in. by passing through rolls. Flaked meats are cooked or tempered prior to pressing or extraction. Animal fats are recovered from fatty tissue by the process of rendering.

Rendering by the dry process is analogous to frying. Well-hashed stock is heated and stirred in open or closed steam-jacketed kettles at about 230°F. The fat is released as protein is coagulated. The cracklings are strained from the fat and pressed to recover the animal fat. Leaf lard is usually dry rendered.

Wet rendering is carried out by cooking in steam autoclaves at pressures of 40–60 psi for 4–6 hours and drawing off the floating fat which is settled or centrifuged.

Pressing of oilseeds is usually done in continuous screw presses. The cracked or flaked, cooked or tempered meats are fed to the screw, which exerts a high pressure and presses the oil out through bars making up a cylindrical cage or barrel. Oil content of the cake is reduced to about 3–4%. Prepressing requires less power, and 10–15% oil is left in the cake and subsequently extracted with solvent.

Extraction of flaked oilseeds or prepress cake with commercial hexane is the most efficient process, removing all but 0.5–1.0% oil. Several types of extractors are used but most operate on a continuous countercurrent principle with the flakes passing through the extractor in the opposite direction from the solvent. A single extractor processes 100 tons or more of soybeans per day or an equivalent quantity of other oil-bearing materials.

Refining is the general term used to describe the over-all process of purifying fats and oils. Specifically it is the initial process of treating the crude oil with alkali to remove free fatty acids, coloring matter, and mucilaginous gums. The oil is mixed with alkali and heated to about 150°F, when it forms soaps of the free fatty acids. Impurities are adsorbed on the soap, which can be settled out but more frequently is continuously separated by centrifuges. Soda ash and sodium hydroxide are refining agents used in one- or two-stage processes. The refined oil is water-washed in centrifuges and dried under vacuum.

Bleaching is an adsorption process for removing color. Refined oil is stirred with about 1.0% of bleaching clay at 220–230°F followed by filtration through a filter press. About 0.1–0.2% of activated carbon may be used with the clay. Bleaching readily removes most of the yellow carotenoid pigments. Special acid-activated clays or carbons are used to remove green pigments present in some oils.

Deodorization is the process of blowing hot oil with steam under a very high vacuum to remove traces of volatile materials causing odor or flavor. Temperatures of 400–475°F and a vacuum of 5–6 mm are usually used in batch, semicontinuous, or continuous processes.

Texturization of plastic fats such as lard or shortening is a rapid solidification process which produces a fine crystal structure and a smooth firm product. The liquefied fats are continuously cooled and solidified on refrigerated chilling rolls or by passage through Votator units. Air or nitrogen is whipped into the product during the process.

Hydrogenation is a catalytic process of converting liquid oils into more-saturated plastic fats by the direct addition of hydrogen. Refined oil and 0.2-0.10% of an active nickel catalyst are charged to a convertor maintained under vacuum. The convertor is equipped with an efficient stirrer for agitation and gas dispersion and a heating jacket. Highly purified hydrogen is introduced under pressure maintained at 5–60 psi. Temperatures range from 200 to 400°F. The composition and physical characteristics of the hydrogenated fat vary with temperature, type and concentration of catalyst, pressure, and agitation or gas dispersion. These must be carefully controlled to give the product desired. The fat is rebleached and filtered after hydrogenation to remove the catalyst.

Crystallization or winterization of cottonseed oil is required for the production of salad oil to prevent solidification or crystallization at refrigerator temperatures. Refined and bleached oil is gradually chilled to about 42°F to crystallize the more-solid glycerides, requiring about 36 hours. Solids are filtered out and the oil deodorized. Soybean oil may be winterized to remove a small amount of wax.

Beef fat is crystallized at 85–90°F to separate oleo oil from oleostearin, and lard can be crystallized at 50°F to separate edible lard oil and grease. Separation is accomplished by pressing.

Emulsifiers, for example, mono- and diglycerides, are added to shortening to improve tenderness and increase shelf life of baked products, particularly bread and cakes. Also used as emulsifiers in margarine, they are prepared by reacting fats with glycerol.

Lecithin is a natural emulsifier recovered from crude soybean oil by water washing. Purified by solvent fractionation, it is widely used as an emulsifier in food products, particularly margarine and chocolate coatings.

Mayonnaise and salad dressings. These are made from salad oils, eggs or egg yolks, vinegar or lemon juice, seasonings, and sugar. Mayonnaise is a semisolid emulsion which must contain 50% oil, although it usually contains more nearly 70% oil. It is prepared by mixing in high-speed beaters.

Salad dressings are similar in composition to mayonnaise but contain a cooked starch base and the oil content is usually 40–50% but may be as low as 20%.

Animal fats. These are produced under federal inspection and are classified as edible or inedible depending upon the source of fatty tissue. They are relatively saturated although they contain a small amount of the highly unsaturated arachidonic acid. Recent advances in processing techniques, particularly interesterification, the rearrangement of glycerides to a random or directed distribution of fatty acids, has greatly increased their utility, and some shortening products are now mixtures of vegetable and animal fats.

Lard and rendered pork fat are produced domestically to the extent of about 2,500,000,000 lb annually. Lard does not of necessity require processing aside from texturization. However, it may receive mild processing treatment to improve flavor and the addition of an antioxidant to improve stability. Iodine values range from 53 to 77 and melting point from 33 to 46°C.

Beef fat or tallow is quite hard, having an iodine value of 40–48 and melting point of 40–47°C. Only limited quantities find use in edible fat products.

Chicken fat has an iodine value of 64–76 and melting point of 32–34°C. Very small quantities are produced and consumption is limited almost entirely to those of the Jewish faith, who use it as a cooking fat.

Vegetable oils. Vegetable oils account for well over one-half the total of edible fats produced and consumed in the United States.

Olive oil production in the United States is limited to about 4,000,000 lb. Prized for its flavor, it is consumed in a crude form without processing. It has a high content of oleic acid, an iodine value range of 80–88, and is liquid at refrigerator temperatures.

Corn oil, produced from corn germ, is a product of the corn-milling industry. Annual domestic production is 300,000,000 lb. It is used almost exclusively as a salad or cooking oil, has a high content of linoleic acid and iodine value of 103–128. *See* CORN.

Soybean oil production has grown tremendously since the introduction of soybeans to the United States from the Orient in the 1920s, and today is over 4,000,000,000 lb. It is used principally in hydrogenated products, that is, shortening and margarine. With an iodine value range of 120–141, it contains major proportions of linoleic acid and 6–8% of linolenic acid. *See* SOYBEAN.

Cocoa butter is the fat pressed or extracted from cacao beans; it is not refined. It has the unique property of melting sharply just below body temperature, producing a cooling sensation in the mouth. It contracts upon solidification and the solid breaks with a sharp fracture or snap. Because of its physical properties, resulting from relatively simple glyceride composition, and its compatibility with chocolate liquor, it is used as a coating fat for confections and other foods. It has an iodine value of 35–40 and a melting point of 28–36°C. *See* COCOA POWDER AND CHOCOLATE.

Coconut oil is a highly saturated fat containing mainly lauric acid, 12 carbon atoms, and other saturated fatty acids of 8–16 carbon chain length. It has extremely good stability to oxidation but develops an off flavor upon hydrolysis. Its utility is limited because it is relatively soft and of short plastic range. It is used principally as a cooking oil, for confectionary fats, and in margarine. Iodine value is 7.5–10.5 and melting point 23–26°C.

Cottonseed oil production in the United States averages about 1,500,000,000 lb. Its major fatty acid is linoleic. It is processed for salad oil, shortening, and for use in margarine manufacture. A liquid oil of good stability, its iodine value ranges from 99 to 113. *See* COTTON.

Peanut oil is produced from peanuts surplus to demand for edible peanuts, or of lower grade, to the extent of 150,000,000 lb annually. It is used principally as a cooking oil and has an iodine value of 84–100. *See* PEANUT. [F.G.D.]

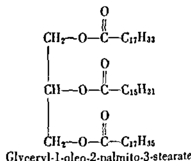
Microbial fat. Production of lipid materials by microorganisms is not economical, except under wartime conditions. Organisms which have been used for this purpose are yeasts, yeastlike organisms that have a primarily respiratory metabolism, and certain molds. Examples of yeasts are *Trichosporon pullulans* (*Endomyces vernalis*), species of *Lipomyces* (one species formerly called *Torulopsis lipofera*), *Candida reukaufii*, *Oospora lactis*, and others. Examples of fungi are *Penicillium javanicum* and *Aspergillus fischeri*. Microorganisms can convert carbohydrates to fat. This process is favored by aerobic growing conditions and under nutrient conditions which limit protein synthesis, that is, when the nitrogen or phosphorus content of the medium is low. Lack of sulfur also causes interference in protein synthesis and increases fat storage. When organisms are grown under such conditions a poor crop of cells is obtained, but the fat content can be 50–60% of the dry weight of the cells. Commercial production is done under compromise conditions, so that a higher cell yield and a somewhat lower fat content are obtained. The lipids include fats, fatty acids, phospholipids, and sterols. Both saturated and unsaturated fatty acids are produced. Extraction of the fats must be preceded by an acid or alkaline hydrolysis of the harvested cells. Ergosterol is the main sterol found in yeast. Strains of *Saccharomyces cerevisiae* have been found which produce 7–10% ergosterol (dry-weight basis). [E.M.M.; H.J.P.]

Bibliography: A. E. Bailey, *Industrial Oil and Fat Products*, 2d ed., 1951; H. J. Deuel, Jr., *The Lipids, Their Chemistry and Biochemistry*, 3 vols., 1951–1957; E. W. Eckey, *Vegetable Fats and Oils*, 1954; *Official and Tentative Methods of the American Oil Chemists' Society*, 1946–1958; S. C. Prescott and C. G. Dunn, *Industrial Microbiology*, 3d ed., 1959

Fat and oil, nonedible

Fats and oils are naturally formed esters of fatty acids with the trihydroxy alcohol, glycerol; they are thus often called triglycerides. Generally speaking, a fat is solid at room temperature, whereas an oil is liquid, a difference that is due either to the molecular size of the fatty acids involved (acids below 10 carbon atoms giving oils), or to the [illegible] FAT AND OIL, EDIBLE.

Structure. Although it is theoretically possible that a fat or triglyceride may be *simple*, *that is*, composed of glycerol esterified three times with the same fatty acid, such is rarely found with natural fats and oils; the general rule is *that the latter are* mixed, meaning that three different fatty acids are involved with one molecule of glycerol. Because many different fatty acids are available (from C_4 to C_{26} in even numbers), both saturated and unsaturated, great variety is encountered in the structures of individual fats or oils. A typical structure for a triglyceride is given by the following (a fat constructed from glycerol and oleic, palmitic, and stearic acids):



Glyceryl-1-oleo-2-palmito-3-stearate

Analysis of fats and oils. A solid fat is relatively or totally free of oil, but the converse is never the case, and in the handling of animal or vegetable oils the first step is chilling, thereby inducing most of the dissolved fat to separate; filtration then follows. The next step is hydrolysis, whereby the ester (fat or oil) is caused to undergo reaction with water to yield glycerol and fatty acids.

Hydrolysis under basic conditions is known as saponification; this has long been an established method for making soaps. Either sodium or potassium hydroxide solution is used, the former giving the sodium salts of the long-chain fatty acids (hard soaps), the latter furnishing potassium salts (soft soaps); in each case, glycerol is the by-product. If the free fatty acids are required, the soaps may be treated with excess mineral acid and the fatty acids collected by filtration, or the fat may be hydrolyzed directly under acid conditions. In this process, superheated steam is led into a vat containing the fat or oil, mixed with a catalyst such as Twitchell's reagent (prepared from benzene or naphthalene, oleic and concentrated sulfuric acids). After hydrolysis is complete, the mixture is chilled and the separated fatty acids are filtered free of the catalyst, water, and glycerol. *See* SOAP AND DETERGENT.

The full chemical characterization of a fat or oil requires the complete identification, qualitative and quantitative, of all the fatty acids present in the substance, but in ordinary commercial work this is not done, the fat or oil being otherwise (crudely) characterized by the determination of various average or mean properties via the following.

Solidification point. The temperature at which the liquid form of the substance changes into the solid form (the solidification point) is not necessarily identical with the melting point.

Acid value (or number). This value is a measure of the amount of free acids present in a body. In the analysis of fats and oils, the acid value is de-

fined as the number of milligrams of potassium hydroxide required to neutralize the free fatty acids present in 1 g of sample.

Saponification value (or number). This is the number of milligrams of potassium hydroxide required for the complete saponification of 1 g of [illegible]

posed of low-molecular-weight, short-chain fatty acids, the value is high. From the saponification value, it is possible to calculate the mean molecular weight of the fat or oil.

Iodine value (or number). The mean unsaturation present in the fatty acids of the fat or oil is expressed as the number of grams of iodine absorbed by 100 g of the sample. The higher the iodine value, the greater will be the unsaturation in the substance.

Reichert-Meissl value (or number). The quantity of low-molecular-weight acids present as esters in the substance is given by the Reichert-Meissl number, defined as the number of milliliters of decinormal alkali solution required for neutralization of the volatile, water-soluble fatty acids obtained from 5.5 g of fat or oil, by following a precise system of saponification, acidification, and distillation. The larger the Reichert-Meissl number, the greater will be the proportion of low-molecular-weight acids present (as esters) in the sample.

The table shows representative, naturally occurring fats and oils, together with their important physical characteristics.

Sources and isolation. Animal fat is stored in many parts of the body, particularly in the superficial fascia under the skin, often forming a layer several inches thick (adipose tissue). In addition, massive amounts are associated with intermuscular

Fats and oils, nonedible[a]

Name and class[b]	Specific gravity, $^{15}/_{15}$°C	Solidification point, °C	Acid value	Saponification value	Iodine value	Reichert-Meissl value
Acorn oil, SV	0.916	−10		199.3	100 0	
Almond oil, NVO	0 914–0.921	−15 to −20	0.5–3.5	183.3–207.6	93–103.4	0.5
Apricot kernel, NVO	0.915–0.926	−17	3.5	191.4–198 2	100–108 7	0.2
Beef marrow, AF	0.931–0.938	29 to 31	1.6	196–199	39–55.4	2
Beef tallow, AF	0 895	31 to 38	0.25	196–200	35 4–42.3	0.25
Black-walnut oil, DV	0.918–0.921	−12	8.6–9.0	190.1–191.5	141–142.7	
Castor oil, NVC	0.960–0.967	−17 to −18	0.12–0.8	175–183	84	1.4
Chaulmoogra oil (USP X Revision), SV	0.950[c]	Below 25		196–213	98–104	
Chinese vegetable tallow, VF	0.918–0.922	24 to 34	2.4	179–206	23–40.5	0.2–0 9
Cocoa (cacao) butter, VF	0.964–0.974	21.5 to 23	1 1–1.9	192.8–195	32.8–41.7	0.3–1.0
Cod-liver oil, MA	0.922–0 931	−3	5 6	171–189	137–166	0.2
Cottonseed oil, SV	0 917–0 918[c]	+12 to −13	0.6–0 9	194–196	103–111.3	0.95
Cottonseed stearin, VF	0.867–0.868[d]	16 to 22	4–10	195	88.7–93.6	0.22
Deer fat, AF	0.962–0.967		0.8–5.3	194.5–200	26–36	0.68
Dolphin oil, MA	0.908–0.930	+5 to −3	2–12	Body 203.4 Jaw 290	Body 126.9 Jaw 32.8	Body 46.9 Jaw 65.9
Horse fat, AF	0.919–0 933	20 to 45	0–2.4	195–200	75–86	1.6–2 1
Human fat, AF	0.9033	15		193–200	57–73	
Linseed oil, DV	0.930–0.938	−19 to −27	1–3.5	188–195	175–202	0 95
Mutton tallow, AF	0.937–0.953	32 to 41	1 7–14	195–196	48–61	
Neat's-foot oil, NA	0.913–0.918	+10 to −2	0.1–0.6	193–199	57.5–75	0.9–1 2
Olive oil, NVO	0.914–0.918	+2 to −6	0 3–1.0	185–196	79–88	0 6–1.5
Palm oil (W. Africa), VF	0.924	35 to 42	10	200–205	49.2–58 9	0 9–1.9
Palm-kernel oil (W. Africa), VF	0.866–0.873[d]			243–255	10.5–17.5	5.0–7.8
Porpoise oil, MA	0.926	−16	Body 1 2 Jaw 5.0	Body 203 4 Jaw 253–272	Body 126.9 Jaw 30.9–49.6	Body 46.9 Jaw 132
Rapeseed oil, NVR	0.913–0.917	−10	0.36–1.0	168–179	94–105	0.0–7.9
Shark oil, MA	0.916–0.919			157–164	115–139	
Soybean oil, SV	0 924–0.927	−10 to −16	0.3–1.8	189–193.5	122–134	0 5–2.8
Sperm oil, Sp	0 878–0.884	15.5	13.2	120–137	80–84	0 6
Tallow (bovine), NA	0 914–0.919	2 to 7.5	0 2–0 25	193 5–199	56–60 5	0 3
Tung oil (China wood), DV	0.939–0 949	Below 17	2	190–197	163–171	1.10
Whale oil, MA	0.917–0.924	−2 to 0	1.9	160–202	90–146	14

[a] Based [illegible]
[b] A[illegible]
anim[illegible]
NVR [illegible]
[c] 25/25°.
[d] 100° C.

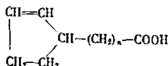
connective tissue and with the heart, kidneys, lungs, and liver. Under constant conditions, the adipose fat of any given species remains relatively constant in chemical nature, but changes in diet and temperature profoundly affect it. Animal fats of commerce are cut away from the carcass during butchering and are purified by heating, which causes them to melt and flow away from protein matter; the heating is either dry (dry-rendering) or by steam

Vegetable fats and oils occur widely throughout the vegetable kingdom; they are found in the leaves, bark, berries, and seeds of plants and trees. The fat or oil is obtained by pressing (hot or cold) in hydraulic presses, or by solvent extraction of the crushed material. The latter method, although expensive, gives the best yield and is coming into general preference.

Some important nonedible fats and oils are listed as follows:

Castor oil, from the seed of the castor-bean plant, is mainly the triglyceride (90%) of (+)-ricinoleic acid, $CH_3(CH_2)_5CHOHCH_2CH{=}CH(CH_2)_7COOH$, and for many years castor oil has been the source of ricinoleic acid used in commerce. Treated with cold sulfuric acid, castor oil undergoes reaction on the hydroxyl group, yielding sulfated castor oil which is known as Turkey Red Oil The latter has long been used to apply the dye alizarin to cloth (mordanted with aluminum salts), to produce the color Turkey Red. If castor oil is dehydrated by the loss of water from the ricinoleic portion of the molecule, a drying oil is produced (*see* DRYING OIL) Long used as a purgative, castor oil has become important as a high-viscosity lubricant for engines.

Chaulmoogra oil, once widely used in the treatment of leprosy, has been shown to consist of the triglycerides of two cyclopentene acids, (+)-hydnocarpic and (+)-chaulmoogric acids, of the following structures:



Hydnocarpic acid, $n = 10$
Chaulmoogric acid, $n = 12$

Olive oil, almond-kernel oil, and neat's-foot oil are the main sources of oleic acid (from 75–80%) of the following (cis) structure, $CH_3(CH_2)_7CH{=}CH(CH_2)_7COOH$.

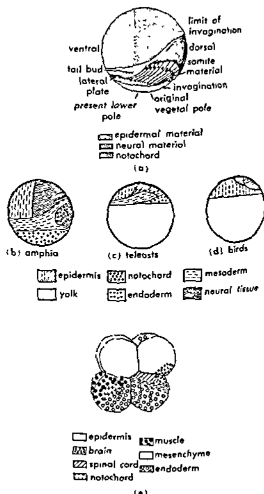
Rapeseed, mustard-seed, and nasturtium-seed oils contain large amounts of erucic acid, $CH_3(CH_2)_7CH{=}CH(CH_2)_{11}COOH$, the acid being cis about the double bond.

Many fish oils (as well as brain tissue) contain the glycerides of nervonic acid, $CH_3(CH_2)_7CH{=}CH(CH_2)_{13}COOH$.

Hydrogenation. Because, in general, the nonedible glycerides of unsaturated acids are oils, whereas those of saturated acids are often edible solid fats, the conversion of the former to the latter is an important commercial process. Thus, cottonseed oil, for example, treated with hydrogen and a nickel catalyst at 200°C, is converted into solid cooking fat. In practice, complete hydrogenation is not desired, because the product is too hard; instead, the oil is hydrogenated only to the proper consistency, control of the latter being maintained through determination of the optimum iodine number. *See* ESSENTIAL OILS; HYDROGENATION; SOLVENT EXTRACTION. [E.B.R.]

Fate maps, embryonic

In development of multicellular animals, the material for a tissue or organ can be traced back to a group of undifferentiated cells or even to a cytoplasmic area of an early developmental stage. The prospective materials for various tissues have definite spatial arrangement at a given developmental



(a) The fate map for the beginning gastrula of a urodele, lateral view (*after W. Vogt, 1926*). (b), (c), and (d) Maps for comparison of distribution of the prospective areas before gastrulation, in three main types of vertebrate development. (e) The fate map for ascidions projected on 8-cell stage. In this map, the individual prospective area is represented by a part of the cytoplasm (*after G. Ortolani, 1953*).

stage. The embryonic fate map is a scheme in which such arrangement is indicated. Introduction of localized vital staining by W. Vogt in 1925 enabled one to pursue a small circumscribed area of the embryo (see illustration, *a*) through the phase of complicated morphogenesis to formation of tissues and organs. This was followed by publication of a number of reliable fate maps for various animal groups. In addition to vital staining, marking with carbon particles or radioisotopes is also used for locating prospective areas. The fate map is not only indispensable for many studies of experimental embryology, but it renders valuable information for the comparative study of embryogenesis of different groups of animals.

In comparing fate maps of amphibians and teleosts at the onset of gastrulation, it is clear that the latter can be derived from the former by increasing the amount of yolk and shifting the whole cellularized area upward (*b* and *c*). Thus in teleosts the invagination of the endoderm and mesoderm occurs at the margin of the blastoderm. On the other hand, in the map of birds, and also in those of reptiles and mammals, the mesodermal areas are surrounded by the ectoderm, so that invagination of the mesoderm occurs within the blastoderm (intradiscal) and not at its margin (*d*). In birds the endoderm is formed underneath the blastoderm, and the mesoderm alone invaginates through the primitive streak. The fate map of the ascidians (*e*) is basically similar to that of the amphibians, except that the prospective muscle is condensed to the ventral half and is not associated with the dorsally located prospective notochord. Thus, the survey of fate maps indicates a general plan of the distribution of the prospective areas common to all groups of chordates. A rather significant change seems to have occurred at this evolutional step, when amniotes appeared with their transdiscal mesoderm. Some fate maps have also been published for the invertebrates and for individual organ rudiments. Operative experiments have demonstrated that prospective areas are able to form tissues other than those corresponding to their actual fate in normal development. This information is, however, not included in the fate map. See CELL LINEAGE; EMBRYOLOGY, EXPERIMENTAL. [T.Y.]

Bibliography: B. H. Willier, P. A. Weiss, and V. Hamburger (eds.), *Analysis of Development*, 1955.

Fault and fault structures

Fractures in rock along which the adjacent rock surfaces are differentially displaced. Some faults are only a few inches long and have displacements measured in fractions of 1 in.; others are miles long with displacements measured in thousands of feet.

Description of faults. The trace of a fault on the earth's surface is a fault line. Its position may be indicated by an escarpment. If the escarpment is the direct result of movement along the fault, it is a fault scarp. On the other hand, if the escarpment is the result of differential erosion of the rocks brought together along the fault, it is a fault-line scarp. In practice the distinction between the two may be difficult to make. See ESCARPMENT.

The surface of a fault may be sinuous and its inclination may change with depth. Where the surface of the fault is other than vertical, the hanging wall is that face of rock which lies above the fault and the foot wall is that which lies beneath. The displacement is not necessarily confined to a single surface but may be distributed over many anastomosing faults.

Differential movement along a fault generally causes granulation and polishing of the adjacent rocks. Gouge is a claylike aggregate produced by the pulverization of the displaced rock. Along some faults the crushed material contains fragments of various sizes, and in this event it is called fault breccia. Mylonite is a microbreccia which is extremely fine-grained and has considerable coherence. It is characterized by a dark color and a streaked or platy structure. Slickensides are polished and striated surfaces that develop along a fault during its dislocation. See MYLONITE.

Nomenclature of displacements. In the absence of precise leveling surveys before and after faulting, the absolute movements on faults cannot be determined. The displacements along faults are

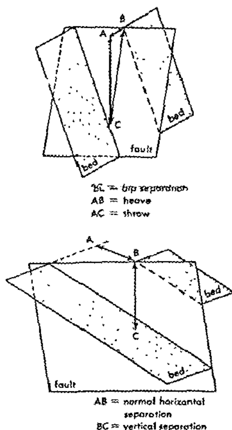


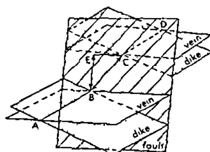
Fig. 1. Disruption of bed by fault.

generally examined in terms of apparent movement or relative movement.

Apparent movement describes the dislocation of beds or other contacts in a given section (Fig. 1). With reference to apparent movement, separation indicates the apparent displacement of two comparable parts of a bed measured in any indicated direction. The normal horizontal separation, or offset, is the separation measured perpendicularly to the disrupted horizon and in a horizontal plane. The vertical separation is the vertical distance between two comparable parts of the displaced bed. The dip separation is the separation measured in a direction directly down the dip of the fault. The throw is the vertical component of the dip separation, and the heave is the horizontal component of the dip separation, both being measured in a section perpendicular to the strike of the fault. Dip is the angle between the maximum slope of a given surface and the horizontal. Strike is the bearing of any horizontal line along a given surface.

Relative movement may be ascertained by combining information concerning the direction of movement on the fault surface with information concerning the apparent movement. In exceptional cases, the amount of relative movement can be determined. For example, the intersection of a vein and a dike, if cut by a fault, would be broken in such a way that the points where the intersection pierced the fault surface would be displaced (Fig. 2). The distance between these points is the net slip. The strike slip is the component of the net slip parallel to the strike of the fault, and the dip slip is the component of the net slip parallel to the dip of the fault.

Classification. Faults may be classified with reference to either apparent or relative movement. With regard to apparent movements, dislocations in which the effects of disrupted contacts can be accounted for by a downward displacement of the hanging wall with respect to the foot wall are normal faults, and those in which the effects of disrupted contacts can be explained by an upward



AB and CD = intersection of dike and vein
BC = net slip
CE = strike slip
BE = dip slip

Fig. 2. Displacement of intersection of dike and vein by fault

movement of the hanging wall with respect to the foot wall are reverse faults. If the relative movement is used as the basis of classification, a gravity fault is one along which the hanging wall has moved down relative to the foot wall, and a thrust fault is one along which the hanging wall has moved up relative to the foot wall. Along a strike-slip fault (wrench fault) the movement is essentially parallel to the strike of the fault.

Mechanics of faulting. According to the Mohr criterion of failure, rupture occurs on planes that include the direction of the intermediate principal stress at values of shearing stress τ determined by some function of the normal stress σ_n.

$$\tau = f(\sigma_n)$$

The shearing stress τ and the normal stress σ_n on these planes are

$$\tau = \frac{\sigma_1 - \sigma_3}{2} \sin 2\alpha$$

$$\sigma_n = \frac{\sigma_1 + \sigma_3}{2} - \frac{\sigma_1 - \sigma_3}{2} \cos 2\alpha$$

where σ_1 is the greatest principal stress, σ_3 is the least principal stress, and α is the angle between the direction of shearing stress and the direction of the greatest principal stress.

The angle α between the direction of greatest principal stress and the plane on which shearing rupture occurs can be calculated by assuming that the value of the shearing stress τ_c necessary to cause rupture is that which is sufficient to overcome a resistance to the occurrence of shearing fracture

$$R = c + \sigma_n \tan \beta$$

where R is resistance to the occurrence of shearing fracture, c is the threshold value of the shearing strength which is a constant for the material, and β is the angle of internal friction. Substituting for σ_n in terms of the principal stresses,

$$R = c + \left(\frac{\sigma_1 + \sigma_3}{2} - \frac{\sigma_1 - \sigma_3}{2} \cos 2\alpha\right) \tan \beta$$

and the difference between resistance to shear and the shearing stress is

$$R - \tau_c = c + \left(\frac{\sigma_1 + \sigma_3}{2} - \frac{\sigma_1 - \sigma_3}{2} \cos 2\alpha\right) \tan \beta - \frac{\sigma_1 - \sigma_3}{2} \sin 2\alpha$$

Rupture occurs on those planes for which the difference between the resistance to shear and the shearing stress is minimal.

$$\frac{d(R - \tau_c)}{d\alpha} = \tan \beta \sin 2\alpha - \cos 2\alpha = 0$$

$$\tan \beta = \cot 2\alpha$$

$$\beta = 90^\circ - 2\alpha$$

$$\alpha = 45^\circ - \frac{\beta}{2}$$

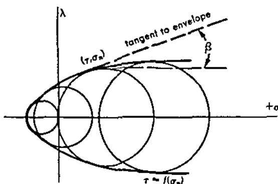


Fig. 3. Mohr's diagram showing the curve $\tau = f(\sigma_n)$ and the angle β.

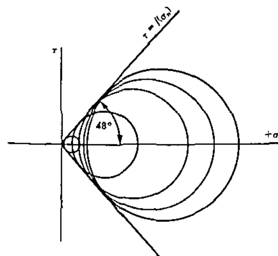


Fig. 4. Mohr's diagram assuming a linear relationship between τ and σ_n and a stress difference of $\sigma_1 = 7\sigma_3$.

The value of β can be found by plotting the curve $\tau = f(\sigma_n)$ graphically, using Mohr's major stress circles. The normal stresses are plotted along the abscissa and the shearing stresses along the ordinate. The envelope of the major stress circles is the curve $\tau = f(\sigma_n)$, and β is the angle between the tangent to the envelope at any coordinates, τ, σ_n, and the abscissa (Fig. 3). Experiments determining failure in cylinders of silicate rocks and minerals under various confining pressures suggest that at failure a linear function exists between the calculated mean stress and the calculated maximum shearing stress. This relation implies that a stress difference of $\sigma_1 = 7\sigma_3$ exists at failure. Using this stress difference to find the envelope of Mohr's major stress circles, β is found to be about 48° and α is therefore about 21° (Fig. 4). Thus, faults in silicate rocks should make angles of about 21° to the maximum principal stress, and the direction of movement on these faults is such that the wedge receiving the greatest compressive force moves inward, whereas the wedge receiving the least compressive force moves outward (Fig. 5).

The geometry of natural faults can be analyzed from the point of view of deriving the orientation of a stress field compatible with the fault pattern. In such a study, a hypothetical segment of the earth's crust is isolated from its surroundings, and certain boundary stresses are imposed on its sides in accordance with Newton's laws of motion. By assuming various values for the boundary stresses, the pattern of the principal stress distribution in the interior of an isotropic block can be established. Possible surfaces of faulting contain the directions in which the intermediate principal stress operates and make an angle of about 21° with the direction in which the greatest principal stress acts (Fig. 6). The set of conditions producing a hypothetical fault pattern identical to that existing in nature serves to establish the stress field causing the natural faults. Analyses of this type assume that the rock is isotropic and that the fault pattern is the result of a single episode of deformation.

Gravity faults. The dips of gravity faults have a maximum frequency of about 60°, suggesting that at the time of faulting the greatest principal stress was oriented approximately vertically and the intermediate and the least principal stresses nearly horizontally (Fig. 7). In many cases, two or more faults parallel one another, commonly with an en echelon pattern. Step faults (Fig. 8) are parallel displacements on which the downthrown side is on the same side of each fault. Under certain circumstances, the blocks between two parallel faults be-

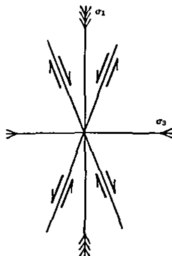


Fig. 5. Displacement on shear planes in relation to directions of principal stresses.



Fig. 6. Directions of principal stresses (solid lines) and of potential fault surfaces (broken lines) compatible with indicated boundary stresses. (From M. K. Hubbert, *Bull. Geol. Soc. Am., 62:335–372, 1951*)

come rotated, and in such cases they are called tilted fault blocks. Gravity faults may be genetically related to local structures, such as folds and domes, or they may be related to broad regional warpings.

Thrust faults. Thrusts have shallow inclinations, the majority having dips between 20 and 30°. Those with dips less than 10° are called overthrusts. If during the course of erosion a part of a thrust mass becomes separated from the main part, the separated remnant is called a klippe. If, on the other hand, erosion cuts a hole through the thrust sheet into the rocks beneath the thrust, the resulting erosional feature is called a fenster. Thrusts are almost entirely confined to regions of folded rocks, and this close association suggests that the two are geneti-

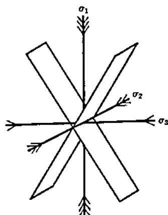


Fig. 7. Orientation of shear planes and principal stresses in normal faulting. σ_1 is nearly vertical. (After E. M. Anderson)

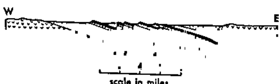


Fig. 8 Step faults and tilted fault blocks of the Connecticut Valley, Connecticut. (*After Barrell, as used in M. P. Billings, Structural Geology, 2d ed., Prentice-Hall, 1954*)

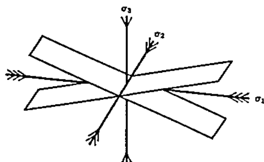


Fig. 9. Orientation of shear planes and principal stresses in thrust faulting. σ_3 is nearly vertical. (After E. M. Anderson)



Fig. 10. The piling up of the Upper and Middle Helvetian sheets in the marginal trough of the Alps, north of the Aar massif. (*After Heim, as used in L. U. De Sitter, Structural Geology, McGraw-Hill, 1956*)

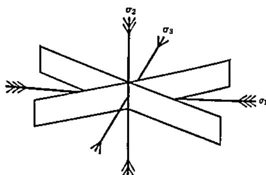


Fig 11. Orientation of shear planes and principal stresses in strike-slip faulting. σ_2 is nearly vertical. (After E. M. Anderson)

cally related. The orientations of these structures indicate that at the time of their development the greatest and intermediate principal stresses were nearly horizontal, and the least principal stress was vertical (Fig. 9).

Certain thrusts are abnormal in that their thrust sheets appear to have moved downhill and to have no visible source (Fig. 10). Some of these thrusts are attributed to gravitational gliding away from the crests of adjacent anticlinal structures.

Strike-slip faults. Strike-slip faults are nearly vertical and commonly are characterized by well-developed shear zones. The orientation of these faults suggests that during their formation the greatest and least principal stresses were approximately horizontal and the intermediate principal stress was vertical (Fig. 11). Although some strike-slip faults may be related to the structures they cut, others clearly postdate the associated structures. Some of the largest strike-slip faults are considered to belong to the major tectonic pattern of the earth, and thus are thought to be the result of deforming pressures that are operative at great depths. *See* TECTONIC PATTERNS; *see also* GRABEN; HORST; [P.H O]

...amics of ...ations to ... distributions and faulting. *Bull. Geol. Soc. Am.*, 62:373–398, 1951; M. K. Hubbert, Mechanical basis for certain familiar geologic structures, *Bull. Geol. Soc. Am.*, 62:355–372, 1951; A. Nadai, *Theory of*

Flow and Fracture of Solids, vol. 1, 2d ed., 1950; E. C. Robertson, Experimental study of the strength of rocks, *Bull. Geol. Soc. Am.*, 66:1275–1314, 1955.

Feather (bird)

A highly specialized, unique, keratinous outgrowth of the skin of birds; one of the most remarkably complex and multiform of organic structures. Collectively the feathers of an individual form a nonconducting body covering (plumage) of light weight and great tensile strength, perfect in design for meeting requirements for aerial life.

Adult feather. A typical adult feather (Fig. 1*a*) consists of a tapering, axial shaft, the rhachis, bearing on each side numerous small, parallel, closely spaced barbs which bear many smaller, barely visible branches, the barbules, on both outer and inner edges. Barbules of adjacent barbs overlap and interlock by means of microscopic hooked barbicels (Fig. 1*b*), to form the coherent, membranous, feather vane. The basal portion of the vane, overlapped by other feathers, is usually fluffy. The barbs and barbules of the basal portion are filamentous and lack hooklets. The base of the feather which is inserted in the skin in a follicle is the calamus; it bears no barbs. At the junction of the calamus and rhachis there is often on the underside a fluffy appendage, the after feather (Fig. 3*d*). Its degree of development is highly variable. Feathers from various body regions of the same and of different species show great variation in each of the above described structures.

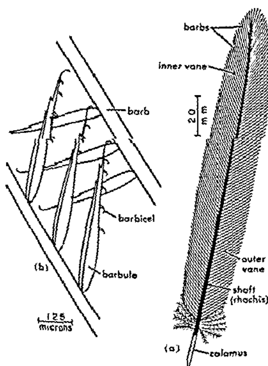


Fig. 1. Typical contour feather (wing). (*a*) Entire feather. (*b*) Barb.

Types. Feathers of the adult plumage, the teleoptiles, are of three main types: contour, down or plumules, and filoplumes (Fig. 3*b-d*). Contour feathers form the bulk of the juvenile and adult plumages. They are visible externally and outline the general form of the body, including wings and tail. The large feathers of the wing, the remiges, and tail, the retrices, are the most highly specialized. Contour feathers arise from certain well-defined areas or tracts, the pterylae, separated by featherless spaces, apteria. The uniform feather distribution of penguins is exceptional. Pterylae vary markedly in form among species. Many peculiar feathers such as the bristles about the mouth and eyes of nocturnal birds, various ornamental plumes, tufts, crests, and others are modified contour feathers. Filoplumes are inconspicuous, degenerate feathers bearing only a few barbs at the tip of a slender hairlike rhachis. They are arranged about the bases of contour feathers, often in groups, and usually covered by them. Down feathers are the simplest and most uniform of adult feathers and closely resemble the nestling down. The long, delicate barbs do not form coherent vanes. Usually concealed by contour feathers, down may occur in either or both apteria and pterylae, or be absent altogether. In ducks, geese, gulls, and similar birds it forms a dense, padlike undercoating. Powder down is a modification in which the tips of the barbs and barbules are brittle and disintegrate into a fine powder as in the herons.

Nestling down feathers or neossoptiles comprise the first soft fluffy plumage found in varying quantity on most newly hatched birds. They are the smallest, most primitive feathers, consisting of a circlet of delicate tapering barbs attached to a short calamus. The barbules are minute and filiform (Fig. 3*a*).

Development. Feathers begin to form early in embryonic life (Fig. 2), in fowl about the end of the first week of incubation. They arise in a definite time-space sequence in groups which correspond to the feather tracts of the adult. The first indication of feather origin occurs in the dermis, the inner layer of skin, as a localized thickening which soon induces a proliferative, feather-forming reaction in the overlying epidermis. Rapid growth of the dermal aggregate thrusts the epidermis upward, to form a domelike elevation (Fig. 2*a*). Each feather germ thus established grows rapidly outward by addition of cells at its base, to form a tapering epidermal cylinder surrounding a dermal core or pulp containing blood vessels which provide nutrients (Fig. 2*b*, *c*). After attaining a certain length, the wall of the epidermal cylinder becomes divided into a series of parallel ridges, the rudiments of the barbs and barbules (Fig. 2*d*). The base of the epidermal cylinder gradually sinks below the skin surface in a tubelike follicle. During its formation, the feather is enclosed by a sheath which ruptures as the chick dries after hatching. With the completion of the down feather, a mass

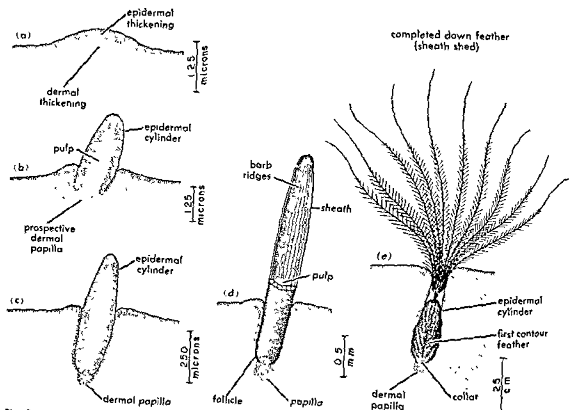


Fig. 2 Development of down feather (diagrammatic). (a) Dermal thickening (b, c, d) Formation of epidermal cylinder (e) Down feather.

of dermal cells covered by a layer of epidermal cells remains in the bottom of the follicle as a permanent structure, the feather papilla (Fig. 2e). It is necessary for the formation of another feather and all succeeding feathers develop from it. As a rule new papillae are not formed after hatching.

The first set of contour feathers begins to form from the papilla before the nestling down is shed (Fig. 2e), but thereafter a new feather does not begin to regenerate until the preceding feather is molted or plucked from the follicle. The permanent papilla is not continuously active but undergoes a resting period after each feather is completed.

When the papilla begins to regenerate a new contour feather, the epidermal cells in contact with the dermal component multiply to form a thick ring of embryonic cells, the collar. Rapid proliferation of cells from its apical margin soon produces an epidermal cylinder enclosing a core, the pulp, derived from the dermal cells of the papilla. All parts of the feather differentiate from the thickened walls of the epidermal cylinder, as a series of close-set parallel ridges (Fig. 4). The shaft forms along the dorsal or outer wall of the cylinder parallel to its axis. The barb ridges arise more or less opposite the shaft, on each side of a ventral locus, parallel to each other and perpendicular to the collar (Fig. 4a). As the barb ridges increase in length by the addition of cells from the collar, they gradually shift to more dorsal positions ultimately joining and fusing with the elongating shaft (Fig. 4a). The feather tip forms first and hence is oldest.

are completed, the feather cylinder gradually emerges from the follicle mouth (Fig. 4b). The pulp is withdrawn toward the base and the surrounding sheath is shed, allowing the feather parts to unfold. The finished feather is a completely keratinized, nonliving structure.

Molting. Throughout life, plumage is renewed periodically. Old, worn feathers are dropped and new ones regenerate from the permanent papilla. Adults undergo a complete molt at least once a year, commonly after the breeding season. This is the postnuptial molt. Additional molts, often partial, may occur also. The special decorative feathers donned by males during breeding seasons are usually the result of partial molts. Young birds undergo a series of molts, complete and partial, before reaching maturity. With each, the individual feathers become progressively more like corresponding ones of the adult in size, structure, and color pattern.

Molting is usually gradual and orderly. Flight feathers, with few exceptions such as the ducks, rails, and geese, are shed and replaced in definite sequence synchronously on the two sides so that the power of flight is not lost. Molting is a complex

physiological process imposing severe strain on the bird's vitality because of the tremendous *amount of blood necessary for development of the* thousands of new feathers.

Feather color. In its extraordinary variety and brilliance, feather color is due to pigments, optical effects produced by the fine structural detail of the minute feather parts, and combinations of the two.

Melanins. These are the most common feather pigments and are responsible for the dark, sober colors such as black, gray, brown, and related tints. They may be uniformly distributed or form elaborate patterns. Melanins are usually present in feathers displaying vivid structural color; they form the groundwork of coloration. The pigment occurs in the form of granules of specific size and shape manufactured by melanocytes, specialized branched cells which develop from colorless precursor cells, or melanoblasts, which originate in the neural crest, a structure located on each side of the neural tube of the early embryo. They migrate to the skin and become incorporated into the developing feather germs where they gradually differentiate into large branched melanocytes filled with pigment granules. After depositing their granules in the feather-forming cells, melanocytes degenerate. The completed feather thus contains innumerable granules of melanin deposited by numerous melanocytes at different time intervals during its formation.

Lipochromes. These are bright, diffuse (nongranular) pigments that birds obtain from their food. They become dissolved in the fats and are transported by the body fluids into the living feather-forming cells. As the feather parts become keratinized, the fat solvent disappears, leaving the

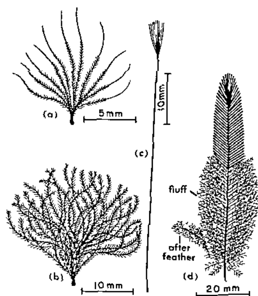


Fig 3. Types of feathers. (a) Nestling down (chick). (b) Down (adult goose) (c) Filoplume (fowl). (d) Contour (back of fowl).

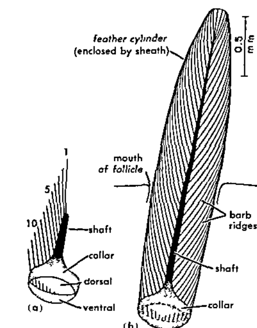


Fig 4. Formation of contour feather (diagrammatic). (a) Early stage. (b) Later stage.

pigment distributed in the cell walls. Reds of the cardinal and flamingo and most bright yellows of the canary and oriole are lipochromes. Mixtures of the two give the orange color that is seen in finches.

Structural colors. These are produced by special modifications of cells of the barbs and barbules of contour feathers. The bright blue of the jay and other blue birds is produced solely by a colorless layer of boxlike cells located in the barbs beneath a thin, transparent, colorless cuticle The walls of the box cells are thick and perforated with extraordinarily fine, canal-like openings filled with air (microvacuoles) which reflect only the blue light waves. Other waves are absorbed by dark melanin in underlying cells. Such blue reflecting cells occur also in barbs of green feathers in parrots and macaws, but the otherwise blue color is changed to green by the presence of yellow lipochrome in the cuticle overlying the blue reflecting cells Red lipochrome in the cuticle and reddish brown melanin beneath the blue reflecting cells produces violet.

Iridescent colors. Iridescent colors of the hummingbird, peacock, pigeon, and similar birds are caused by the interference of light waves reflected

water. The colors change as the observer's position changes with reference to the light source. All colors of the rainbow may be represented with green being common. The presence of a dark pigment (melanin) is necessary for the full manifestation of iridescence and is responsible for metallic luster and brilliance. [M.E.R.]

Feces

The waste material eliminated by the gastrointestinal tract. The components of feces vary greatly, depending upon diet, metabolic activity, and other factors. In general, the feces contain undigested food materials and the excretions or secretions added by the digestive system. Most food passes into the lower bowel in a semiliquid state; water is then reabsorbed and the feces become more solid. In any case, water is normally a large component, along with a high content of bacteria which are saprophytic occupants of the intestine. Inorganic salts, pigments, bile products, and cellular debris are present. Illness, changes in diet or water intake and excretion, emotional conditions, and many drugs and chemicals may alter the feces both qualitatively and quantitatively. *See* DIGESTIVE SYSTEM. [E.C ST.]

Feed water

A relatively pure condensate or treated water, usually a mixture of the two, used to supply a steam generating unit. The proportions vary according to the amount of make-up required to compensate for losses from the system supplied by the unit. In a closed power generating cycle, these losses may be less than 1% and justify the cost of demineralization or evaporation to maintain high purity. When steam is used for industrial process consumption, the return of condensate may be partially or completely impractical, thus requiring larger proportions of treated water make-up, in some cases equal to the entire feed water demand. Make-up water is usually processed chemically. (*See* RAW WATER.)

Three distinct sources of feed water are natural or raw water available to the plant, feed water as supplied to the unit, and boiler water contained in the generating system.

Natural waters always contain impurities, to some degree, in the form of suspended or dissolved solid matter, and dissolved gases, which originate from earlier contact with earth and atmosphere. Contamination may also occur from industrial or municipal wastes. These substances, when concentrated by recirculation in the generating system, may attack metals corrosively, or form insulating deposits of sediment or scale on heat transfer surfaces, which result in overheating and possible failure of pressure parts.

The principal scale-forming constituents normally encountered are compounds of calcium and magnesium, or silica. The principal corrosive agents are dissolved oxygen and carbon dioxide. Oil and grease impair wetting and heat removal from the generating surfaces and may also form corrosive scale or sludge.

Feed water is treated to remove or alter impurities so that they will not interfere with the steam plant. The principal treatment consists of removing dissolved oxygen and other gases before admitting the water to the boiler. Almost all gases can be driven out by boiling the water at slightly above atmospheric pressure in open heaters and discharging the noncondensable gases through a vent condenser.

More thorough separation is obtainable from spray or tray-type deaerating heaters, arranged for countercurrent scavenging of released gases to prevent their re-solution.

In high purity systems, attention must be given to preventing corrosion of piping and equipment in the preboiler system, by maintaining an alkaline condition of the condensate feed water through the use of ammonia or volatile amines added to the boiler or steam lines for mixture with the steam.

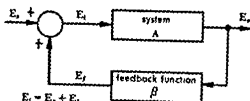
The need for and type of water treatment varies considerably for different plants, depending on the nature of the water supply and the percentage of condensate available as feed water. It becomes more imperative and critical for units operated at high pressures, with high rates of heat transfer. *See* BOILER FEED WATER; STEAM GENERATING UNIT; WATER TREATMENT. [F.G.E.]

Feedback circuit

A circuit that returns a portion of the output signal of an electronic circuit or control system to the input of the circuit or system. When the signal returned (the feedback signal) is at the same phase as the input signal, the feedback is called positive or regenerative. When the feedback signal is of opposite phase to that of the input signal, the feedback is negative or degenerative.

The use of feedback in electronic circuits and automatic control systems produces changes in the characteristics of the system which improve the performance of the system. In electronic circuits, feedback is employed either to alter the shape of the frequency-response characteristics of an amplifier circuit and thereby produce more uniform amplification over a range of frequencies, or to produce conditions for oscillation in an oscillator circuit. In automatic control systems, feedback is used to compare the actual output of a system with a desired output, the difference being used as the input signal to a controller. These two points of view will be considered separately in the following discussion. However, the analysis of both feedback amplifiers and electromechanical control systems can be made on a frequency response basis; from the point of view of analysis the two have much in common.

Amplifier feedback. Feedback can be introduced into an amplifier by connecting to the input a fraction of the voltage appearing across the output terminals. An amplifier will, in general, have better frequency-response characteristics when the system has feedback than when there is no feedback. The system can be designed to have a wider band width and more nearly ideal frequency-response characteristics. Furthermore, harmonic distortion caused by nonlinear tube characteristics can be reduced by the use of feedback. *See* DISTORTION (ELECTRONIC CIRCUITS). The use of feedback in an improperly designed system, however, can produce



Block diagram of feedback circuit.

a system with worse characteristics. Amplifiers can become oscillators when feedback is used in an improperly designed system. For a discussion of amplifier frequency response and bandwidth *see* AMPLIFIER.

A system with feedback can be analyzed by using the block diagram representation shown in the illustration. The sinusoidal input signal is E_s and the amplifier gain is A, which is a function of frequency. When there is no feedback $E_o = AE_s$ because $E_i = E_s$. When there is feedback, $E_i = E_s + E_f$.

Since $E_o = AE_i$, and $E_f = \beta E_o$, the over-all gain of the system with feedback is then

$$\frac{E_o}{E_s} = \frac{A}{1 - A\beta}$$

This formula for the gain of an amplifier with feedback indicates the effect of feedback upon the frequency response of the amplifier without feedback. The $(1 - A\beta)$ term in the denominator is a complex number. Therefore, the magnitude and phase angle of the gain of the amplifier with feedback will differ from the gain of the amplifier without feedback. The amount of the difference depends upon the value of the A and the β terms, and no general statements can be made.

If there is a frequency for which $A\beta = 1$, the denominator of the expression for the gain will be zero while the numerator will not be zero. When this occurs, the amplifier will oscillate at approximately the frequency for which $A\beta = 1$. Furthermore, if there is a frequency for which the magnitude of $A\beta$ is greater than unity and the phase angle of $A\beta$ is 0°, 360°, or any integral multiple of 360°, the amplifier will oscillate.

Positive and negative feedback. The terms positive feedback and negative feedback are used to denote the type of feedback found in certain electronic circuits.

When the feedback factor β is negative, the denominator of the feedback equation will be greater than unity, and the over-all gain with this negative feedback will be less than the gain with no feedback. As long as the product $A\beta$ is negative, stable operation will result.

The product $A\beta$ becomes positive for a positive feedback factor β. If $A\beta$ is positive but less than unity, the operation is stable and increased gain results. However, an increase in $A\beta$ to unity throws the system into oscillation.

It follows that both the magnitude and the phase angle θ of the product $A\beta$ are important in determining whether the feedback is positive or negative relative to the input voltage.

The phase angle θ is a real function of frequency and is not a constant. Therefore, the voltage fed back cannot be exactly out of phase with the input signal except at one frequency (or possibly at several frequencies). At these frequencies the feedback is negative. In fact, it is possible for the phase angle θ to be zero or an even multiple of 180° at frequencies within the frequency range of operation. Thus there are frequencies for which the feedback signal is in phase with the input signal. At these frequencies the feedback is positive.

In the usual resistance-capacitance-coupled amplifier stage there is a range of frequencies for which the phase angle of the output voltage with respect to the input voltage is 180°. This frequency range lies within the somewhat larger range called the midband. If a given amplifier composed of several stages has an input signal frequency such that each stage has the 180° phase relation between input and output, then negative feedback can be introduced around an odd number of stages. A multistage amplifier may have several feedback paths.

The effect of feedback upon the frequency response of an amplifier can be determined from the expression $A/(1 - A\beta)$. The performance in a particular case depends upon the behavior of A and β as functions of frequency.

The usual method for analyzing the feedback amplifier to determine the possibilities of oscillation is the examination of the amplitude and phase of $A\beta$ over the frequency range from zero to a sufficiently high frequency (on the order of 100,000 cps). The examination is often made by plotting the magnitude of $A\beta$ against the phase angle in polar coordinates. The plot shows the magnitude of $A\beta$ when the phase angle is 180°. If it is greater than or equal to one, the amplifier will oscillate.

An alternate method of analysis is to plot the logarithm of the magnitude of $A\beta$ against the logarithm of frequency and the phase angle against the logarithm of frequency. The value of frequency at which the phase angle equals 180° is determined from the phase-angle-vs.-log-frequency curve. If for this value of frequency the logarithm of $|A\beta|$ is greater than zero (the magnitude of $A\beta$ is greater than one), the amplifier will oscillate.

Several practical problems are found in the analysis of feedback amplifiers. One is that the values of the various circuit components, such as resistors, capacitors, tubes, and transformers, are only approximately known. For example, resistors may have a 20% tolerance and the values for tube parameters may have a 20% variation around a nominal value. Another problem is that all of the parameters which affect the performance, in particular stray capacitance introduced by the physical layout of components and wiring, are not included in the expression for $A\beta$. This means that, in practice, mathematical analysis is only a guide; final design is refined by experiment.

The discussion has considered voltage feedback where the output voltage, or some function of it, is fed back to the input. Current feedback, where a

voltage proportional to the output current or some function of the output current is fed back, may also be used. Furthermore, an amplifier may have both voltage and current feedback.

Oscillator feedback. An oscillator can be viewed

From this viewpoint the condition for oscillation at a frequency f_o is (to a first and usually very good approximation) that $A\beta = 1$ at $f = f_o$. This means that the feedback must be positive. However, most oscillator circuits are sufficiently simple to make it easier to determine the conditions for oscillation directly from the analysis of the circuit rather than by trying to determine A and β. *See* OSCILLATOR.

Servomechanism feedback. The purpose of feedback in a servomechanism is to obtain better control of the system. As an example, consider a position control system which is used to position an object at a point determined by a reference signal. The command to move the object is derived by comparing the reference signal with a signal indicating the instantaneous location of the object. The command signal is an error signal derived from the comparison of the actual and desired signals; any error signal drives the system in such a direction as to reduce the error, that is, to make the actual position agree with the desired reference position.

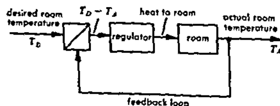
If feedback were not used in the position control system, a precisely calibrated control device would be needed to position the object in each position dictated by the reference. In general the required control could not be built with sufficient accuracy. *See* CONTROL SYSTEMS; SERVOMECHANISM. [H.F.K.]

Bibliography: J. D. Ryder, *Engineering Electronics*, 1957, J. G. Thomason, *Linear Feedback Analysis*, 1955.

Feedback control system

A system in which the value of some output quantity is controlled by feeding back the value of the controlled quantity and using it to manipulate an input quantity in such a manner as to bring the value of the controlled quantity closer to a desired value. Such a system is also called a closed-loop control system.

The illustration shows a schematic picture of a



Regulated temperature control system.

supplied to the room is changed by the regulator in a measured manner which will tend to make this difference disappear. Such feedback systems are widely used in electronic amplifiers, in the control of pressure and flow in chemical processes, and in the guidance and control of airplanes, missiles, and space vehicles. For a further discussion of feedback control systems *see* CONTROL SYSTEMS. [J.A.H.]

Feeding mechanisms and digestion

Various mechanisms have evolved in the different animal groups which have enabled these organisms to utilize a wide range of food materials. Modifications in feeding mechanisms are associated with the available food in the organism's particular habitat. The form and function of the digestive tract in the different groups of animals can usually be correlated with both the type of food utilized and the feeding mechanism involved.

FOOD

Animals are dependent for their supply of energy and for the growth and maintenance of their tissues upon an intake of complex organic materials. It is these, together with vitamins and mineral salts, which constitute their food.

Soluble food. A considerable amount of dissolved organic matter is always present in natural waters; the sea, for example, contains at least 6 milligrams per liter. At one time it was thought that animals might be able to utilize this organic matter by absorbing it through their body surface, but probably only bacteria and perhaps a few Protozoa such as the dinoflagellates can obtain much nutriment in this way. Marine fish, however, are known to swallow sea water, and there is evidence that some fresh-water fish do also. Dissolved organic material might be absorbed from such ingested water, but it seems doubtful that this could provide more than a very small fraction of the nutritional requirements of an animal.

The only easily demonstrable examples of animals relying upon soluble food are those in which this method of feeding is part of a highly specialized mode of life. The parasitic tapeworms and Acanthocephala have, for example, lost their alimentary canal and feed by absorbing through their body wall the soluble materials which have been produced by the digestive activity of their hosts. Many parasitic protozoa feed in a similar way. Those trypanosomes which parasitize the blood stream of vertebrates obtain their energy by absorbing the circulating blood sugar of the host. Certain free-living animals also rely upon liquid food. Familiar examples are provided by the many insects (bees, ants, and mosquitoes) which feed upon the body fluids or secretions of plants and animals, and by the Arachnida (scorpions, spiders, and mites) which feed on fluid material from the bodies of their living or dead prey.

Microphagy. This term is applied to the process of feeding upon small particles. Large quantities of small particles are suspended in water, either as

minute living plants and animals or as products of the excretion or decay of these or larger organisms. Animals such as *Cardium* (Fig. 3*a*) which utilize them are termed suspension feeders and are found in many groups of the animal kingdom, including the Coelenterata, polychaete worms, arthropods, mollusks, and protochordates. These show a remarkable diversity in the details of their feeding mechanisms, but common features are setting up currents in the water by movement of cilia or appendages, trapping particles in a sticky secretion of mucus, and complex filtering and rejection mechanisms which are characteristic of the filter feeders.

The vibration of the thoracic limbs of the common water flea *Daphnia* (Fig. 1*a*) draws under the carapace a current of water which is filtered by the bristles (setae) on these limbs (Fig. 1*b*). Large particles are rejected by this filter, but small ones can pass on into a food groove from which they are transferred into the mouth with the aid of a sticky secretion. Filter feeding is also convenient for sedentary or sessile animals. The permanently sessile barnacle *Balanus*, for example, creates a current of water by vibrating its limbs when they are protruded through the valves of its protective shell.

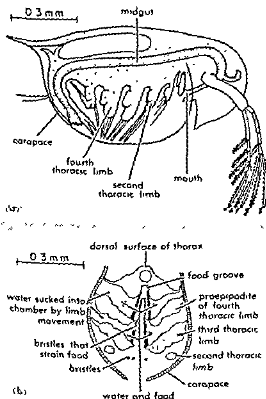


Fig. 1. (a) *Daphnia* (*Simplified after W. Bullough, Practical Invertebrate Anatomy, Macmillan, 1950*). (b) Diagram of transverse section through *Daphnia* to indicate the mode of filter feeding (*After Storch, from L. A. Borradaile et al., The Invertebrata, 3d ed. rev., Cambridge, 1958*).

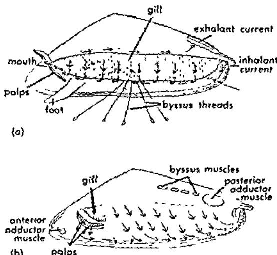


Fig. 2. Diagrams to show the ciliary currents employed in the filter feeding of *Mytilus*. (a) Food currents, with the left lobe of the mantle removed. (b) Rejection currents, with foot and gills removed to show the interior of the right lobe of the mantle. (*Adapted from Orton, from L. A. Borradaile et al., The Invertebrata, 3d ed. rev., Cambridge, 1958*)

The use of cilia in filter feeding is seen in lamellibranch mollusks such as the mussel *Mytilus*, which creates a current of water with the cilia on its gills. The water passes down an inhalant siphon into the mantle cavity (Fig. 2). Here the heavier particles tend to drop onto the mantle wall because the water flow rate is reduced. They are swept into the exhalant siphon by ciliary action. The remaining particles are caught on certain of the cilia of the gill filaments and then, entangled in mucus, are driven by ciliary tracts down to the lower edge of the gill and forward toward the mouth. Labial palps surrounding the mouth serve for further filtering and rejection; heavier particles are again discarded onto the mantle wall while lighter ones are carried into the mouth.

Considerable amounts of particulate food collect on the substratum at the bottom of the water as detritus. Animals using this for food are referred to as detritus or deposit feeders. The distinction between these and the suspension feeders described above is not, however, a sharp one, and the principles involved are quite similar. Detritus-feeding bivalves such as *Tellina* (Fig. 3*b*) have long flexible siphons in contrast to the short relatively inflexible ones of its suspension-feeding relatives. The siphons protrude from the sand or mud, and the food material is drawn into the gill chamber by ciliary action. Another detritus feeder is the sand-burrowing sea urchin, *Echinocardium* (Fig. 3*c*), which has numerous tube-feet with suckerlike extremities. In some genera these are used exclusively for locomotion, but in this particular genus some feet are greatly elongated and can be extended from the mouth of the burrow where they pick up the detritus particles and transfer them to shorter tube-feet near the mouth (see ECHINODERMATA). Cilia co-

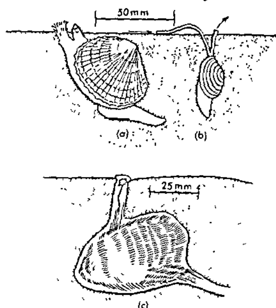


Fig. 3. Sand-burrowing bivalves showing positions in sand during feeding. (a) Cockle, *Cardium edule*, a suspension feeder with short siphons. (b) *Tellina tenuis*, a deposit feeder with longer and more flexible siphons. (c) Sand-burrowing urchin, *Echinocardium cordatum*, within sand, to show one of the tube-feet extended to the surface for feeding. (From C. M. Yonge, *The Sea Shore*, Collins, 1949)

ing the body surface create a current of water into the burrow, but in this case the current is primarily for respiration rather than for feeding, although it does serve also to sweep away waste material.

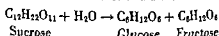
Macrophagy. A microphagous animal needs to filter enormous quantities of water in order to secure adequate supplies of food; it is calculated, for example, that a 2-year-old mussel filters 60 liters of water per day. This tends to restrict the size of such animals, because metabolism (and hence food requirements) varies in proportion to more than the square of the linear dimensions. Microphagy is normally practicable only for aquatic forms, although earthworms and their marine equivalents, lugworms, constitute a partial exception to this by being able to utilize the minute food particles present in soil and mud through swallowing large quantities of such material (see ANNELIDA).

A fundamentally important step in evolution has been the development in various groups of animals of macrophagy, in which large masses of plant or animal food are seized. Typical examples are the larger Crustacea (crabs and lobsters), the cephalopod Mollusca (squids and cuttlefish) and the vertebrates, but even quite lowly animals have been able to exploit large prey by virtue of specialized adaptations. A striking example of this is the Portuguese man-of-war (*Physalia*), essentially a floating colony of coelenterate polyps and medusae, which is able by means of its tentacles and stinging threads to trap and feed upon relatively large fish.

The development of macrophagy by the higher groups of animals has led to greatly increased size of the individual and has opened up the possibility of developing new modes of life, including the exploitation of the diversity of habitats provided by life on dry land. It has also necessarily involved specialization of the feeding mechanism along lines different from those characteristic of microphagy. Examples of this are the tentacles, beaks, and poison glands of cephalopods; and the buccal cavity of higher vertebrates which with its associated jaws, teeth, tongue, and sensory taste buds is of fundamental importance in the selection and seizing of the food and often in the breaking of it into smaller particles in preparation for subsequent digestive action. Some other aspects of macrophagy will be referred to below.

DIGESTION

The complex molecules of organic food must be broken down into their structural units before they can be assimilated into the body tissues. This breakdown, which is termed digestion, involves hydrolysis, or the addition of water to the molecules. Hydrolysis is represented in the following equation for the production of two molecules of monosaccharides from one molecule of a disaccharide



$$\underset{\text{Sucrose}}{C_{12}H_{22}O_{11}} + H_2O \rightarrow \underset{\text{Glucose}}{C_6H_{12}O_6} + \underset{\text{Fructose}}{C_6H_{12}O_6}$$

Digestive enzymes. For such reactions to take place under the conditions available within the body it is necessary for them to be catalyzed by hydrolytic digestive enzymes. These share with other enzymes certain properties which greatly influence the organization of the alimentary canal, since a primary function of the latter is to produce adequate amounts of enzymes in the right place at the right time. Thus, the digestive enzymes are highly specific in their action, each acting upon a particular type of molecule or even upon a particular type of bond within that molecule. Their speed of enzymatic action increases with temperature to a maximum at about 40°C, above which the speed diminishes and the enzymes are finally destroyed. This property would appear to favor the birds and mammals which maintain their bodies at a constant temperature close to this optimum rather than the lower vertebrates whose temperature fluctuates with that of the environment. However, like other proteins, enzymes appear to show some variation from species to species, and there is evidence that they may be adapted to the conditions in which they have to work. For example, the pepsin of the Pacific king salmon differs from that of the pig in being less responsive to temperature changes. Another important property of enzymes is the dependence of their speed of action upon the pH or hydrogen ion concentration of the medium in which they act. Each has an optimum pH value at which its activity is maximal. See THERMOREGULATION.

Proteases. These enzymes act upon the peptide bonds of protein molecules, and are classified into

exopeptidases and endopeptidases. The former act only upon the terminal bonds of the protein chain, while the latter can act in addition upon certain of the more centrally situated bonds. Well-known examples of endopeptidases are pepsin with a pH optimum of 1.5 to 2.5, and trypsin with an optimum between 7 and 9. Chymotrypsin resembles trypsin and has a powerful milk-clotting action, while rennin also clots milk at an optimum pH of 6.0 to 6.5. Examples of exopeptidases are carboxypeptidases, so-called because they act on terminal bonds associated with unsubstituted carboxyl groups; aminopeptidases, acting where there is a free amino group; and dipeptidases, acting on bonds uniting two amino acid residues into a dipeptide molecule.

Carbohydrases. These highly specific enzymes hydrolyze carbohydrates into their monosaccharide units. For example, amylase, which is activated by chloride ions, breaks down starch into the disaccharide maltose, while maltase breaks down maltose into the monosaccharide glucose. Invertase hydrolyzes sucrose into invert sugar consisting of equal proportions of glucose and fructose, while lactase breaks down lactose into galactose and glucose.

Lipases. The fats of food, which consist largely of triglycerides of the higher fatty acids, are hydrolyzed by lipases into their constituent molecules of glycerol and fatty acids.

Digestive enzymes and adaptation. The enzymes listed previously (with the exception of pepsin) are widely distributed throughout the animal kingdom, but their relative activity in different species shows clear evidence of their adaptation to the animal's diet. Carnivores such as coelenterates, starfish, and squids have powerful proteases and weak carbohydrases and corals are quite unable to digest starch, while herbivores such as most bivalve mollusks and tunicates have weak proteases and strong amylases.

Further evidence of adaptation is found in certain species which possess unusual enzymes that are clearly related to unusual diets. A striking example is the remarkable ability of the snail (*Helix*) and the shipworm (*Teredo*) to digest cellulose by means of a cellulase. This enzyme enables the latter animal to feed exclusively upon wood which it grinds into sawdust with the valves of its shell. The wood-eating termites, however, lack a cellulase of their own and depend for their nutrition on the digestion of wood by Protozoa which live in their hindgut (see HYPERMASTIGIDA; ISOPTERA). Mammalian herbivores also lack this enzyme, despite the importance of cellulose in their diet, and they too make use of microorganisms The complex stomach of ruminants (cattle and sheep) functions as a fermentation chamber in which bacteria and Protozoa break down the plant food with the production of large quantities of carbon dioxide, methane, and volatile acids, mainly acetic and butyric. The chewing of the cud aids in the completion of this process (see DIGESTIVE SYSTEM). In the horse and rabbit similar fermentation is carried on in the enlarged cecum. These animals are not ruminants but the digestion of cellulose is aided in the rabbit by this animal's habit of eating its feces. Another example of unusual digestive adaptation is the presence of chitinase in *Helix* and in the larva of the ectoparasitic wasp *Pseudogenia*, which discharges this enzyme externally for the digestion of the integument of its host spider. There is some evidence for the occurrence of amygdalase, salicinase, and lichenase in certain herbivorous fish. Further research might discover other examples of such adaptations which need not depend upon the presence of unusual enzymes. For example, the ability of the clothes moth larva to digest the keratin of wool, a substance not attacked by the usual proteases, seems to depend primarily upon the fact that a strong reducing agent in its alimentary canal breaks the disulfide cross-linkages upon which the great stability of this particular substance depends. When the peptide chains are opened in this way, the intestinal protease is able to act. This enzyme shows an interesting adaptation in that, unlike most proteases, it is not inhibited by the SH groups which result from the reduction of the disulfide bonds.

Intracellular and extracellular digestion. Digestion must take place in a cavity in which the enzymes can be brought into contact with the food. This principle is seen in its simplest form in many of the noncellular Protozoa. In *Amoeba*, for example, food particles are engulfed into a food vacuole. Within this vacuole they can be seen to disintegrate gradually, presumably under the influence of enzymes which are secreted into the vacuole. They diminish in amount as the soluble products are absorbed into the protoplasm, and undigestible residues are eventually discarded through the body surface. The contents of the vacuole show an acid reaction (perhaps as low as pH 3.6) at first, but this is a result of the respiration and death of the ingested organisms and is not due to the secretion of acid. Later it approximates neutrality (pH 7.3) and then provides a suitable medium for the action of enzymes which are able to attack protein and fat, but there is some doubt as to whether *Amoeba* can digest starch.

Digestion which takes place in this way within the cytoplasm of the organism is termed intracellular digestion. With the evolution of multicellular organisms it became possible for digestion to take place outside the cells in a cavity called the alimentary canal. The food is taken into the cavity and then mixed with enzymes which are secreted into it from the cells of the digestive epithelium which forms the wall of the canal. The products of digestion are absorbed through the epithelium into the tissues of the body. This is termed extracellular digestion. It is seen in its simplest form in the coelenterates in which the alimentary canal (coelenteron) has a single opening through which food enters and undigested residues leave. However, in these animals, such as *Hydra*, much use is still made of intracellular digestion. It is probable that

only the preliminary digestion of proteins is truly extracellular with the remainder of the digestion taking place in vacuoles in the cells of the digestive epithelium, food particles being ingested by these in much the same way as in *Amoeba.* Intracellular digestion persists in many other Metazoa, partly as a primitive feature in simpler animals (Porifera, Ctenophora, Turbellaria) but also as an adaptation particularly associated with microphagy in more advanced forms (Mollusca, *Amphioxus*). The assumption of macrophagy inevitably leads to increasing emphasis on extracellular digestion, but it is doubtful whether the intracellular method is ever entirely lost, for there is evidence that in fish, and even in mammals, the final stages of protein digestion are completed by intracellular enzymes There may also be an intracellular lipase in fish. Intracellular digestion may be effected by wandering phagocytes; for example, in the stomach of bivalve mollusks they appear to ingest some of the larger particles which are unable to enter the digestive gland.

ALIMENTARY CANAL OF VERTEBRATES

Important factors in the organization of the alimentary canal of higher animals are the establishment of functionally distinct regions and the increase in surface area of the alimentary epithelium, a special aspect of the latter being the outgrowth of the epithelium to form glands which remain connected with the main canal by ducts. These features are seen in the vertebrates. In man (Fig. 4), for example, the food is taken first into the buccal cavity, where it is chewed and mixed with saliva containing an amylase which initiates the digestion of starch. It then passes through the pharynx into the esophagus, from which point onward it is propelled by peristaltic contraction of involuntary muscle in the wall of the canal and is no longer under conscious control. Movement of the food is aided by the production of mucus throughout the length of the alimentary canal. Mucus is a slimy substance, commonly rich in acid mucopolysaccharides, and is produced by cells distributed throughout the lining epithelium of the digestive tract. It is because of the abundance of these cells that this lining is termed the alimentary mucosa.

Stomach. The first of the two main phases of digestion takes place in the stomach in which the food may be retained for up to 4 hours by the closure of a muscle sphincter. It is churned and mixed with the gastric juice, produced in part by the lining mucosa, which secretes mucus, and in part by the branched tubular glands which are developed by folding and ingrowth of that mucosa These glands contain several different types of cell which give rise to more mucus, to hydrochloric acid, and to the enzyme pepsin. The acid provides an optimum pH of about 1.5 for the action of the pepsin. Rennin is also present, and perhaps a small amount of lipase, but the main function of the stomach is to break down proteins into peptones and proteases.

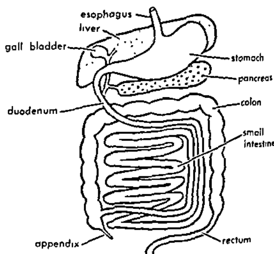


Fig. 4. Diagram of the human alimentary tract.

Intestine, pancreas, and liver. The gastric contents (chyme) pass at intervals into the duodenum, which is the anterior region of the small intestine. This is followed by the large intestine (colon) and rectum. The large intestine bears a small blind cecum and appendix. Three secretions are discharged into the small intestine. Two are produced by the pancreas and liver, glands which have grown as evaginations from the wall of the alimentary tract, and the third by the intestine itself. Together they partially neutralize the acidity of the chyme and establish a pH of 5.5 to 6.8.

The pancreatic secretion is an alkaline fluid containing trypsinogen, an inactive substance which is activated by enterokinase in the intestinal secretion, succus entericus, to form the active enzyme trypsin which continues the digestion of proteins Also present are the enzymes chymotrypsin, carboxypeptidase, amylase, and lipase.

The secretion of the liver is known as bile and is released after concentration and temporary storage in the gall bladder. It contains no digestive enzymes. Its importance in digestion results mainly from its content of bile salts, the sodium salts of glycocholic acid and taurocholic acid. In addition to creating an alkaline reaction in the bile, these substances activate lipase and, because of their property of lowering surface tension, help to emulsify the fats in the food and so make it easier for the lipase to act on them. It is also thought that they may aid in the absorption of the products of fat digestion by forming water-soluble compounds with the fatty acids. The succus entericus arises mainly from the crypts of Lieberkühn, simple tubular glands formed by ingrowth of the duodenal mucosa. An alkaline secretion is also contributed by Brünner's glands, situated at the upper end of the duodenum and more complex in their form. The presence of aminopeptidase and dipeptidase provides for the completion of protein digestion, probably in part within the mucosal cells, while invertase, maltase, and lactase complete the digestion of carbohydrates. Thus as the food passes through

the alimentary tract, it is acted on by a sequence or chain of enzymes, the proteins and carbohydrates being broken down in a series of well defined steps.

Absorption. Amino acids and monosaccharides pass into the blood stream, but fats are resynthesized and collect in special lymph vessels (lacteals). The digestion products in man are mainly absorbed in the small intestine, the mucosal surface of which is increased by the outgrowth of innumerable fingerlike villi.

Absorption is now known to be a process of considerable complexity which cannot be accounted for by simple diffusion through the mucosa. For example, xylose has a smaller molecule than glucose, but the latter is much more readily absorbed than is the former, a phenomenon which must be ascribed to active uptake by the mucosal cells, resulting in selective absorption and the expenditure of energy by the cells concerned. The possible role of bile salts in the absorption of fatty acids has been mentioned. There is evidence also that unhydrolyzed fat may be taken up by the cells, provided that it has been emulsified into sufficiently small droplets.

Water is mainly absorbed in the large intestine in man, and abundant secretion of mucus takes place here in order to lubricate the passage of the feces. This region is also important in the excretion of insoluble waste material which cannot be discharged through the kidney.

Evolutionary aspects. The acid phase of digestion in the stomach seems to be a peculiarity of vertebrates and to be a later addition to their digestive mechanism. The lampreys and hagfish, survivors of the ancient group of primitively jawless vertebrates (Agnatha), have no stomach. There is good reason for believing this condition to be a direct inheritance from the filter-feeding protochordates which were their immediate ancestors. The modern Agnatha feed by sucking, and it seems likely that the stomach did not evolve until the appearance of jaws made macrophagous feeding possible and created the need for a region of the alimentary tract in which bulky prey could be lodged. The secretion of acid may have been an adaptation for the softening and disinfection of the food (the gastric mucus and acid of mammals are known to have antiseptic properties), with pepsin appearing as a still later and consequential adaptation. That gastric digestion is not an essential phase, however, is shown by the fact that the stomach has been secondarily lost in various fish, including the Holocephali, the Dipnoi (lungfish), and a number of teleosts such as the carp. It appears also to be more simply organized in the lower vertebrates, for the acid and pepsinogen are produced there by the same cell instead of by separate ones as in mammals.

It is possible to trace something of the evolutionary history of other features of the vertebrate alimentary tract. For example, the pancreas is not yet developed in the Agnatha but is represented in lampreys by cells which are concentrated in the mucosa at the anterior end of the intestine and which are particularly concerned with the secretion of trypsin. Increase of surface area in lampreys and in many fish is provided for by the development of an internal ridge, the typhlosole, a device similar in principle to that employed by such invertebrates as earthworms. Later in evolution, in some fish and in all higher vertebrates, this ridge is lost and is replaced by a more complex folding and lengthening of the intestine, which becomes increasingly coiled to be accommodated within the limited space afforded by the body cavity.

Control of digestive activity. It is clear that if the alimentary system is to operate with maximum efficiency there is need for close control and coordination of the movement of the food material and of the discharge of the secretions. The study of this aspect in mammals, the only group in which it has been at all thoroughly analyzed, has shown that there is a remarkable combination of nervous and secretory mechanisms involved. It is well established that the alimentary canal is under the control of the autonomic nervous system, operating through its two components, the sympathetic and the parasympathetic systems. As a very general rule, to which there are many exceptions in detail, sympathetic stimulation tends to arrest movement and secretory activity while parasympa-

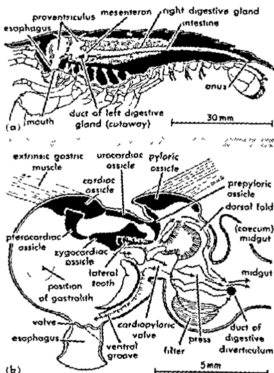


Fig. 5 (a) Crayfish, *Astacus*, opened from the left side to show the alimentary canal. (b) Esophagus and proventriculus of crayfish. Large arrows show the route of solid particles, broken line and arrows the route of the digestive fluid (From A. J. Grove and G. E. Newell, *Animal Biology*, Univ. Tutorial Press, 1955)

thetic stimulation has the opposite effect. Examples of such nervous control are the inhibition of alimentary activity during periods of pain and anger, the secretion of saliva in response to the sight of food, and the initiation of gastric secretion (the so-called psychic secretion) in a hungry dog when it is shown meat but before it has begun to eat it.

The best established illustration of control by a secretory mechanism is the release of secretin from the wall of the duodenum into the blood stream when the acid gastric contents are released through the pyloric sphincter. Secretin is a hormone which circulates in the blood. It reaches the pancreas and stimulates that organ to release pancreatic fluid which arrives in the duodenum exactly when it is needed. There are other examples of digestive hormones, although the evidence for their existence is less complete. Gastrin is believed to stimulate gastric secretion and to be released from the wall of the stomach when food enters this organ and enterogastrone is believed to inhibit gastric secretion and to be released from the intestine in response to a fatty diet. Some parts of the alimentary tract, notably the [illegible]
to [illegible]
m[illegible]
by [illegible]
with other aspects of alimentary organization it is likely to be the result of the gradual development of complexity during evolutionary history.

OTHER TYPES OF ALIMENTARY CANAL

While the alimentary canals of some invertebrates are anatomically quite simple (Fig. 1), others are at least as complex as those of vertebrates, but it is impossible to make more than brief reference to them here. An understanding of their functioning has often been impeded by the drawing of unfounded comparisons between their parts and those of the mammalian system, for, although in describing them it may be convenient to use mammalian terminology, it must be remembered that each group has its own alimentary organization which has evolved in adaptation to its own feeding habits and mode of life.

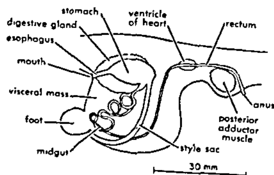


Fig. 6 Diagram of alimentory canal of a bivalve, the clam *Mya*, from the left side. (From C. M. Yonge, *Studies of the comparative physiology of digestion*, *J. Exptl. Biol.*, 1:15–63, 1923)

For example, in crayfish and lobsters (Fig. 5*a*) the seizing and biting of food is a function of certain external appendages, including the mandibles. After entering the mouth the food is passed through a short esophagus (Fig. 5*b*) into a proventriculus ("stomach"). The anterior region of this, the gizzard, is provided with specialized musculature and calcareous plates which form the gastric mill. This grinds the food and mixes it with the digestive secretion which passes forward from the digestive gland in which it is formed. Digestion takes place in the proventriculus; the fluid digestion products are filtered through hairs in a posterior filtering chamber of this organ and reach the midgut from which they enter the digestive gland. The latter has often been called the liver or hepato-pancreas, but both terms are misleading since this gland differs from both the pancreas and the liver of vertebrates in being formed of delicate tubules through which the food and secretions are circulated by muscular action, absorption as well as secretion taking place within them. Indigestible solid particles follow different routes backward through the filtering chamber, and the relatively long hindgut conducts them to the anus.

The specialized requirements of microphagous feeders are seen in the bivalve mollusks (Fig. 6) in which, in contrast to mammals and arthropods, the movement of food through the alimentary tract depends largely upon cilia instead of muscle. Particles are carried down the esophagus into a stomach. A rod, the crystalline style, projects into this from a style sac and is kept in continuous rotation by ciliary action, its head being slowly worn away as a result of friction with a horny gastric shield. The effect of this is to entangle and wind the mucus threads with their contained food particles, and to mix them with an extracellular amylase which is released from the style. The smaller particles are sorted out and pass into the tubules of a digestive gland, in which much intracellular digestion occurs. Most of the larger particles are passed back into the intestine, although some may be ingested by phagocytes in the stomach. Similar principles are seen at work in the microphagous protochordate *Amphioxus*, in which rotation, mixing, and sorting of food material and secretions takes place in the midgut by ciliary action; the smaller particles enter a blind digestive cecum where intracellular digestion occurs.

[E.J.W.B.]

Bibliography: M. E. Brown (ed.), *The Physiology of Fishes*, vol. 1, 1957; C. L. Prosser (ed.), *Comparative Animal Physiology*, 1950.

Feldspar

A group of silicate minerals that make up about 60% of the outer 15 km of the earth's crust. The feldspars constitute the most abundant group of minerals and are also important economic minerals, particularly in the ceramic and glass industries which utilize most of the output as raw materials. Feldspars are also used as gemstones if nicely col-

ored or distinguished by an attractive luster. *See* SILICATE MINERALS.

Chemical composition. The feldspars are silicates of aluminum, Al, with the metals potassium, K, sodium, Na, and calcium, Ca, and rarely barium, Ba.

Their chemical composition can usually be expressed as a sum $Or_xAb_yAn_z$ with $x + y + z = 100$ and the symbols having the following significance: Or, $KAlSi_3O_8$; Ab, $NaAlSi_3O_8$; and An, $CaAl_2Si_2O_8$. The symbols Or, Ab, and An are derived from the special feldspar species: orthoclase, albite, and anorthite. Rubidium, Rb, and strontium, Sr, may enter the feldspars in amounts usually below 1 mole % of $RbAlSi_3O_8$ or $SrAl_2Si_2O_8$. Larger amounts of Ba may be found as $BaAl_2Si_2O_8$ in potassium feldspars (which are then called hyalophane). A pure barium feldspar, $BaAl_2Si_2O_8$ (celsian), is very rare. Al can be replaced by Fe (rarely more than 5%). All intermediate members, Or-Ab and Ab-An, are known. Whereas Or does not take much An into solid solution or vice versa, Ab can take up both Or and An (up to a maximum composition of about $Or_{20}Ab_{70}An_{10}$).

Occurrence. The feldspars occur as components of all kinds of rocks. In igneous and metamorphic rocks their particle size usually lies between 0.1 and 10 mm; in porphyritic rocks it frequently reaches 5–10 cm; in pegmatites, crystals of 10 m or larger are found. The size of authigenic (rather pure potassium or sodium) feldspars, developed in sedimentary rocks, is usually below 1 mm. Feldspars are also found beautifully crystallized, as fissure minerals in clefts or as druse minerals in cavities, for example in granites.

Properties. The properties of the feldspars are variable and depend upon their chemical composition, the conditions under which they grew, and on their subsequent history (changes of the external conditions with time).

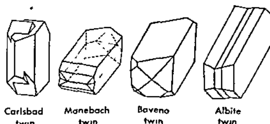
Optical properties. In agreement with their symmetry (triclinic or monoclinic) they are birefringent (relatively low, approximately 0.01 and smaller) with refractive indices about 1.52 (potassium feldspar), 1.53 (sodium feldspar), and 1.58 (calcium feldspar). They can be optically negative or positive corresponding to their chemical composition, crystal structure, and state of homogeneity. *See* CRYSTAL OPTICS.

Color. When feldspars appear in the state in which they grew, that is, as homogeneous crystals, they are usually clear and translucent. However, they have usually undergone changes in the course of geological times (exsolutions, changes in symmetry with subsequent small-scale twinning, and metasomatic alterations with or without deviation from a possible feldspar composition) as a result of which they have lost their translucency and assumed a milky or dirty appearance. Feldspars have no color of their own but are frequently colored (yellow, brown, reddish, or dirty green to black) by impurities. The cause of the beautiful bright green color of amazonite is still unknown. Crystals more or less exsolved on a submicroscopic scale may show a whitish-bluish luster (moonstones are alkali feldspars or cryptoperthites), a blue luster (peristerite is an oligoclase), or a bright luster of all colors (labradorites). Aventurine exhibits a shiny red or green color as a result of small but microscopically visible exsolved hematite or included mica particles.

Mechanical properties. The hardness (Mohs scale 6) is somewhat lower than that of quartz (Mohs scale 7). Feldspars are brittle but usually show an excellent cleavage parallel to the (001) plane and another, somewhat less distinct, parallel to the (010) plane. In addition (and sometimes even more pronounced than the cleavages), there are parting planes parallel to (110), $(1\bar{1}0)$, and other directions. These occur especially in clear, unmixed plagioclases that may not show the normal cleavage behavior at all.

Morphology. Feldspars are monoclinic, $\alpha = \gamma = 90°$; $\beta \approx 116°$, or triclinic (α and γ deviating from 90° by up to about 4°). The ratio $a{:}b{:}c$ is about 8.6 A:13 0 A:7.2 A for potassium feldspar and 8.1 A:12.8 A:7.2 A for sodium and calcium feldspar. Depending upon chemical composition and growth conditions their habit may be isometric, elongated after the *a*, *b*, or *c* axis or tabular after (010) or (001). The main limiting faces (110), $(1\bar{1}0)$ are always present in well-developed crystals, and (010), (001), $(\bar{1}01)$, $(\bar{2}01)$. If (110) and $(1\bar{1}0)$ are the dominating faces (as in adularia and in anorthoclase of the rhomb porphyries), a double-wedge-shaped morphology results with rhomblike cross sections.

Twinning. Twinning is frequently present in feldspars. A distinction can be made between simple and multiple twinning. The former is characteristic of a few individuals related to each other by one or more of the following twin laws (only the most common are listed here): (1) after (100) or [001], Karlsbad law; (2) after (021), Baveno law; (3) after (001), Manebach law; (4) after (010), albite law; and (5) after [010], pericline law. The twin laws (4) and (5) cannot, for reasons of symmetry, occur in monoclinic feldspars. In addition triclinic feldspars untwinned or twinned in the above simple sense frequently show multiple twinning, that is, many alternating lamellae whose orientation is governed throughout by the same



Twinning in feldspars. (*From C S. Hurlbut, Jr., Dana's Manual of Mineralogy, 16th ed., Wiley, 1952*).

twin law. The twin laws that most frequently lead to multiple twinning are the albite and the pericline laws. The width of the lamellae in multiple twins of plagioclases is usually about 1 mm and smaller; in the alkali feldspars anorthoclase and orthoclase, the width is frequently below the range of microscopic visibility. In such cases the existence of twinning can be determined only by x-rays. If the multiple twinning is governed by the albite law, the composition plane of the lamellae is (010). In the case of pericline law the composition plane is the so-called rhombic section, an irrational plane in the zone [010], the orientation of which depends on the lattice geometry. If the angle between the rhombic section and (001) is called $+s$ and conforms to the definition that the $+a$-axis direction corresponds to 0° and the $+c$-axis direction to 116°, then this angle s is expressed by the relation $\tan s = \cos \gamma \cdot \tan \alpha^*$, where α^* is the angle $(001) \wedge (010)$. Frequently the orientation of the rhombic section is expressed by the value of σ, an angle between the trace of the rhombic section on (010) and the a-axis. σ is expressed by the relation $\cot \sigma = \tan \gamma \cdot \cos \alpha^*$. In feldspars the difference between the values s and σ is small (usually less than 4°). Because γ varies considerably in the plagioclases as a function of Al/Si distribution and chemical composition, s and σ are rather important as diagnostic magnitudes. *See* TWINNING (CRYSTALLOGRAPHY)

Microcline-type twinning. Because the alkali feldspars and the acid plagioclases can be monoclinic or triclinic, a transition from monoclinic to triclinic may be connected with simultaneous (sometimes submicroscopic) twinning after the albite and pericline laws with the twin plane (010) perpendicular to the twin axis [010]. Thus, a multiple twin system is formed, consisting of numerous interpenetrating domains divided into four groups of orientations: A_1 and A_2 (with albite-twin orientation) and P_1 and P_2 (with pericline-twin orientation).

Genetically, three main types of twinning can be recognized: (1) growth twins, usually observable as simple twinning according to all the laws mentioned; (2) transformation twins, usually observable as an interpenetrating system of multiple twinning after the albite and pericline laws; and (3) deformation twins (also called mechanical twins), usually observable as multiple twinning after the albite, or pericline laws, or both. Because, in principle, multiple twinning after a single law may [illegible]
or d[illegible]
une[illegible]
pretation of such twins. In the case of microcline-type twinning, however, it can be rigidly concluded that such material originally grew with monoclinic symmetry.

Melting point. Alkali feldspars melt (potassium feldspar incongruently, forming a mixture of leucite and glass), at about 1100°C. Plagioclases melt (as a function of An content) between 1100 and 1511°C.

Crystal structure. Although the crystal structure of the feldspars has been known in principle since 1933 (W. H. Taylor), details are still the object of investigation in several laboratories. Each Si^{4+} (radius ≈ 0.4 Å) and Al^{3+} (radius ≈ 0.5 Å) is surrounded in a nearly ideal tetrahedral coordination by four O^{2-} (radius ≈ 1.3 Å). Thus, each O^{2-} belongs to two neighboring SiO_4 or AlO_4 tetrahedrons. Considering the Si^{4+} or Al^{3+} alone, they form a three-dimensional network with the coordination number 4 and the O^{2-} (nearly) centering the shortest distances between them.

This $(Si,Al)O_2$ framework has sizable voids in which the larger cations K^+ (radius ≈ 1.3), Na^+ (radius ≈ 1.0), and Ca^{2+} (radius ≈ 1.0 Å) are located. In the solid solutions, the K, Na, and Ca ions always appear to be randomly distributed. On the other hand, the distribution of the Si and Al ions may differ largely for feldspars of the same chemical composition. The ordered or disordered state of Si/Al is in a great measure responsible for the different feldspar varieties and their varying properties.

In discussing the various main possibilities connected with the Al/Si–order/disorder, it must be mentioned that the feldspar unit cell contains 4 molecules $(K,Na,Ca)(Si,Al)_4O_8$. There are, therefore, 16 small ions (Si or Al) per cell. In the case of monoclinic symmetry these 16 ions occupy two unequivalent 8-fold point positions that may be called A and B. Restricting the discussion at first to the alkali feldspars, $KAlSi_3O_8$ and $NaAlSi_3O_8$, two extremes are theoretically possible: (1) 6Si + 2Al are situated in A, and 6Si + 2Al are situated in B, (2) 8Si are situated in A and 4Si + 4Al are situated in B. In both cases some disorder or random distribution occurs, in case (1) more so than in (2). States intermediate between (1) and (2) are possible. A higher state of order can only be reached if the symmetry sinks from monoclinic into triclinic. Then the two 8-fold point positions A and B split up into four 4-fold point positions A_1, A_2, B_1, and B_2, and the 4 Al can completely fill one of these, for example B_2, whereas A_1, A_2, and B_1 are occupied solely by the 12 Si. Such arrangement, case (3), would be the most ordered one compatible with the feldspar-type structure. Whereas case (3) must necessarily be triclinic, it must be realized that in the above-mentioned cases (1) and (2) the symmetry may be monoclinic, but is not necessarily so. Summarizing this discussion, the table lists those cases of varying order/disorder and varying symmetry that appear to be pertinent for the understanding of the alkali feldspar varieties now to be discussed in detail. (In the plagioclases, see below, additional cases have to be discussed as additional order/disorder states are present because the Al/Si ratio varies from 1:3 in $NaAlSi_3O_8$ to 1:1 in $CaAl_2Si_2O_8$.) *See* CRYSTAL STRUCTURE; CRYSTALLOGRAPHY.

The existence of a high degree of disorder in sanidine and analbite and a high degree of order in microcline and albite has recently been proved by nuclear magnetic resonance experiments (E. Brun-

St. Hafner, P. Hartman, F. Laves, H. H. Staub, 1960). *See* MAGNETIC RESONANCE.

ALKALI FELDSPARS

Alkali feldspars is the name for feldspars composed of Or and Ab in any ratio. Usually some An is present as a minor constituent. The amount of An that can be taken up by the alkali feldspars increases with the Ab/Or ratio. The maximum value of An/(Or + Ab) that is found in natural crystals is approximately 1% in Or-rich feldspars and may increase to approximately 15% in Ab-rich feldspars. Because $NaAlSi_3O_8$ (Ab) is the Na-rich member of the alkali feldspars as well as of the plagioclases, Ab feldspars are sometimes considered to belong to the former, sometimes to the latter. It appears reasonable to separate the alkali feldspars from the plagioclases by the Or/An ratio (in alkali feldspars Or/An > 1; in plagioclases Or/An < 1). There are many names for the different varieties of alkali feldspars, most of which are listed and characterized below.

Sanidine. This is the monoclinic modification of $KAlSi_3O_8$. Its optical and structural properties vary with the Ab content and the temperature of equilibrium. At high temperatures the Al/Si distribution appears to be that of case (1) in the table. At lower temperature the Al/Si distribution changes in the direction of case (1–2) mcl. Corresponding to this change a subdivision into sanidine (high, intermediate, and low) appears justified. Connected with these structural changes are those of the optical properties. In Na-free sanidine (high) the optical plane is parallel to (010) with 2V ≈ 60°. With the change (high → intermediate → low), 2V becomes smaller, may pass through 0° and again increase, but now within an optical plane perpendicular (010) and inclined about 10° to (001). The maximum value of 2V in the plane perpendicular to (010), is not yet known, because there is a continuous transition of the properties of sanidine into those of orthoclase. Within the optical plane (010) 2V diminishes not only with increasing Al/Si order but also with increasing Ab content. A very useful parameter easily measurable by an x-ray method, for determining the Ab content is the decrease with increasing Na content of the a-lattice parameter (~8.6 for potassium feldspar, ~8.1 for sodium feldspar). Sanidine appears to be stable under equilibrium conditions above approximately 500°C. Below this temperature microcline becomes stable (the temperature is not yet exactly known).

Microcline. Microcline is the triclinic modification of potassium feldspar. Its Al/Si distribution is ordered in such a way that monoclinic symmetry is from a structural point of view impossible. The triclinic deviation from monoclinic symmetry may be small. In the most ordered state it approaches case (3) tr of the table, having values $\alpha^* = (010)/(001) \approx 90°25'$ and $\gamma^* = (010)/(100) \approx 92°16'$ (maximum microcline). All intermediate states between these values and $\alpha^* = \gamma^* = 90°$ are possible and known from natural occurrences. These intermediate states can be characterized as possessing Al/Si distributions lying between case (1–2) mcl ≈ case (1–2) tr and case (3) tr of the table. These intermediate structural states have optical properties that are also of intermediate character (that is, lying between the triclinic optics of maximum microcline with 2V ≈ 60°, in a plane not quite perpendicular to (010), and the mono-

Main types of Al/Si distribution possible under equilibrium conditions in monoclinic and triclinic alkali feldspars, as a function of temperature

Temperature	Point positions				Symmetry	
	A		B		Monoclinic	Triclinic
~1100°C	6Si + 2Al		6Si + 2Al		Case (1)mcl[a]	Case (1)tr[b]
↑	6–8Si + 2–0Al		6–4Si + 2–4Al		Case (1–2)mcl	Case (1–2)tr[b]
	8Si		4Si + 4Al		Case (2)mcl	Case (2)tr[c]
	A_1	A_2	B_1	B_2		
	3Si + 1Al	3Si + 1Al	3Si + 1Al	3Si + 1Al	Case (1)mcl[d]	Case (1)tr[e]
	3–4Si + 1–0Al		mSi + nAl ($m > n$)	mSi + nAl ($m < n$)		Case (1–3)tr
~0°C	4Si	4Si	4Si	4Al		Case (3)tr

[a] As the positions A and B possess different environments, it is improbable that the Al distribution can be exactly equal in both positions. Therefore, the case (1)mcl can only be a theoretical limit which may, however, be very nearly attained at temperatures near the melting point (about 1100°C).

[b] For reasons of structure theory these cases cannot have distributions exactly as listed. For the monoclinic positions A and B are split up into the triclinic positions A_1, A_2 and B_1, B_2, and it cannot be expected that the Al distribution be exactly the same in A_1 and A_2, as in B_1 and B_2 respectively.

[c] Case (2) is the theoretically possible case of monoclinic symmetry with highest possible order. It shows the direction in which a completely disordered feldspar can develop a more ordered (or less disordered) distribution, while retaining monoclinic symmetry. For thermodynamical reasons only the intermediate states of case (1–2)mcl can be expected to occur until about (7Si + 1Al) are in A and (5Si + 3Al) in B.

[d] This case is only monoclinic in the exact sense if both A_1 and A_2 and B_1 and B_2 are indistinguishable. If this is true, the case is identical with the one listed in the first row of the table.

[e] Considerations similar to those mentioned in footnote b hold for the case (1)tr.

clinic optics of sanidine] Under equilibrium conditions maximum microcline does not take much Ab into solid solution (less than about 2 mole %). Most of the Ab content found in untreated microcline crystals (usually amounting to about 20 mole %) is present as perthitic, or cryptoperthitic exsolution, or both. However, it is possible to prepare the whole series of solid solutions between microcline and albite (see below) under nonequilibrium conditions (superheating) by a heat treatment at approximately 1000°C, that is, above the transformation temperature microcline → sanidine. The time of heat treatment must be long enough to allow a K/Na exchange to take place, but short enough to prevent changes of the Al/Si distribution. Microcline usually shows multiple twinning of the microcline type. From this type of twinning it may be concluded that most microclines originally grew as sanidine, either stably at high temperature or metastably at low temperature. In the course of geological times they underwent a continuous change by diffusive transformation into the triclinic state now present. This transformation involving Al/Si-exchange at fairly low temperatures proceeds very slowly and has not yet been reproduced in the laboratory. The slowness of this ordering process is the reason why all states of microcline are found in nature reaching from monoclinic properties to the triclinic ones of maximum microcline. The domain size of twinned microcline is variable, usually < 1 mm. The domains can be submicroscopic in size and then the material appears to be optically monoclinic. In these cases a distinction of microcline and orthoclase can only be made by means of x-rays. Whereas it appears to be impossible to produce microcline in the laboratory by synthesis or by heat treatment of sanidine below the sanidine microcline transformation temperature (assumed to be about 500°C), the reverse process can be effected within times available for extended experiments. It takes place at about 1000°C or higher in days, weeks, months, or years, depending on factors (impurities, imperfections) still unknown. Bright green colored microclines (the cause of the color is still unknown) are used as gemstones and called amazonite or amazon stone. Such material is always maximum microcline. The reverse does not hold (most maximum microclines [illegible]

[illegible]

ing the appearance of an eutectic growth produced by simultaneous crystallization, have already changed into the microcline structure.

Orthoclase. The name orthoclase is used quite differently by different people. In the widest sense the name applies [illegible]

[illegible] ...optically twinned analbite would be orthoclases or sodium orthoclases if a considerable amount of Na (more than 25 mole % Ab) is found by chemical analysis. Petrographers usually call a material orthoclase, if it appears to be optically monoclinic with the optical plane ⊥(010) and $2V > 30°$ (if $2V < 30°$, such material is frequently called sanidine). As stated above there are true monoclinic and true triclinic states of potassium feldspar, sanidine and microcline. Therefore, it appears practical to include in the term normal orthoclase a material whose monoclinic appearance depends on the twinning of domains having an Al/Si distribution of triclinic symmetry. If these domains are submicroscopic in size or if their Al/Si order is small, such material shows optically perfect or nearly perfect monoclinic character. The same holds for normally exposed x-ray photographs which show reflections that can be given monoclinic indices. Long-exposed x-ray photographs, however, reveal additional diffuse streaks originating from the main reflections. The position of these streaks indicates the presence of triclinic domains within the normal orthoclase with a mutual orientation similar to that of microcline twinning. Thus, the normal orthoclases may display a broad variety of structural features with respect to the size of domains and the degree of Al/Si order within them. This concept also explains the variability of their optical behavior.

Adularia is another variety of potassium feldspar, grown as well-developed crystals in fissures, especially in the region of the Swiss Alps. The predominant forms are (110) and ($\bar{1}$01). The form (001) is frequently very small or absent. The morphology is monoclinic and the structure is in principle similar to that of normal orthoclase. However, thin sections investigated with a polarizing microscope between crossed nicol prisms usually reveal a much more complex behavior than the normal orthoclases. Frequently different parts of the same individual display varying optical properties ranging from those of typical orthoclase to those of typical sanidine. In places they deviate strongly from monoclinic symmetry and are usually most triclinic where the (110) and ($\bar{1}$10) faces or the (1$\bar{1}$0) and ($\bar{1}\bar{1}$0) faces meet. The triclinicity of such parts can be measured by x-rays. Deviations from monoclinic symmetry of 30′ for γ^* are frequently met with in Swiss adularias. A variety from Mexico (valencianite) shows deviations in places extending to those of maximum microcline. The manner in which the optically different parts are related to each other indicates that adularia originally grew metastably as sanidine at low temperatures. In time the variety adularia changed to triclinic symmetry, a process taking place at different speed in the outer and inner parts of the crystals. Different conditions in different localities are the reason why the adularias display a fascinating variety of optical behavior when investigated between crossed nicols. Adularia frequently develops simple (growth) twinning after the Baveno and Manebach laws. The Karlsbad law is rare. The crystals are usually clear, sometimes milky, but

colorless. They may reach sizes up to 10 cm (rarely even more) and are frequently covered by green chlorite.

Albite. $NaAlSi_3O_8$ is chemically the Na-rich end member of the alkali feldspars. If it is ordered with respect to the Al/Si distribution to the same extent as microcline, it is called albite and is the phase of $NaAlSi_3O_8$ most frequently found in nature. It usually occurs in the most highly ordered state, case (3)tr of the table (corresponding to maximum microcline), with a, b, c = 8.16, 12.77, 7.17 and α^*, β, γ^* = 86.3°, 116.7°, 90.5°. Albite occurs in all groups of rocks and is usually colorless or milky white. In fissures it frequently forms beautifully developed glassy-clear crystals, tabular after (010) and mostly simply twinned after the albite law. Sometimes fourlings are found and then two albite-twinned plates are, in addition, twinned after the Karlsbad law. A variety twinned after [010] (pericline law) and elongated along the same direction is called pericline. There are indications, such as the position of the rhombic section, that such material (which always appears milky as a consequence of numerous small voids) did not grow originally as a pure albite phase but as oligoclase which subsequently underwent metasomatic changes in chemical composition. Albite crystals in rocks frequently exhibit polysynthetic twinning, predominantly after the albite law.

Under equilibrium conditions and at low temperatures albite does not take more than 1–2 mole % of Or or An into solid solution. However, [illegible]

cation of $NaAlSi_3O_8$ which under conditions of equilibrium is stable only at high temperature. Its existence and monoclinic symmetry can be proved by x-ray investigations at high temperature and by the twinning behavior. This albite material kept for long periods (weeks to months) at temperatures near the melting point (1000–1100°C) exhibits a displacive, unquenchable transformation into a triclinic form, analbite. If an untwinned single crystal of albite is treated in this way, submicroscopic twins of the microcline type fashioned after both the albite and the pericline laws appear in the course of cooling. Optically such material seems to be monoclinic but its triclinic character can easily be proved by x-rays. Such triclinic $NaAlSi_3O_8$ that upon heating becomes monoclinic by a displacive transformation (that is, by a reversible transformation which takes place at once when the transformation temperature is passed) should be called analbite. The temperature at which the transformation monalbite ↔ analbite takes place, is a function of the Al/Si distribution. Highly disordered monalbite that approximates case (1)mcl of the table transforms into analbite at a lower temperature than does monalbite with less disorder, having an Al/Si distribution similar to case (1–2)mcl of the table. This influence of the Al/Si distribution on the transformation temperature of monalbite ↔ analbite is very impressive. It appears that the highest transformation temperature (strictly speaking the only one possible under equilibrium conditions if the influence of pressure is neglected) is about 900°C. Under nonequilibrium conditions, temperatures as low as −60°C are reported (W. L. Brown, 1959). Thus a monoclinic $NaAlSi_3O_8$ can exist metastably at room temperature, provided it has a very high degree of disorder resulting either from direct growth or assumed during prolonged heat treatment at very high temperatures near the melting point. Monalbite forms a complete series of solid solutions with sanidine. Thus the monoclinic high-temperature modifications of the alkali feldspars may be subdivided into sanidine, sodium sanidine, potassium monalbite, and monalbite. How high an An-content can be taken up by monalbite is still unknown. Present data indicate the existence of calcium-monalbites extending to about $Ab_{80}An_{20}$ at high temperatures. This fact is rather important for a genetic interpretation of the feldspars in rhomb porphyries which have compositions of about $Or_{10}Ab_{80}An_{10}$. The peculiar appearance of their twinning indicates that they originally grew as monoclinic potassium, calcium, or calcium-potassium monalbites that became polysynthetically twinned and unmixed during the course of cooling.

Analbite. Analbite, a triclinic $NaAlSi_3O_8$ that is not stable at any temperature, has an Al/Si distribution lying between cases (1)tr and (1–2)tr of the table, which according to structure theory cannot, strictly speaking, exist stably under any conditions. It can be produced by cooling monalbite and can be changed back to monalbite by heating, provided the Al/Si distribution did not change and assume the character of cases (1–3)tr or (3)tr given in the table. After having undergone such a change, the material could only approach monoclinic symmetry when reheated. It would actually be structurally intermediate between analbite and albite, if albite is defined as a series of materials having Al/Si distributions which necessitate triclinic symmetry and are capable of existing under equilibrium conditions. Lattice constants of most triclinic analbite are about a, b, $c \approx$ 8.12, 12.87, 7.13 and α^*, β, $\gamma^* \approx$ 86.0, 116.5, 88.3. All intermediate values between these and $\alpha^* = \gamma^* = 90°$ (monoclinic symmetry) may be met with as a function of temperature and Al/Si distribution (analbite-monalbite series). In addition intermediate values between those given here and those given above for albite can also be expected (unstable analbite-albite series) as well as those of intermediate states that are stable as a function of temperature under equilibrium conditions forming a series albite (low, intermediate, high)—monalbite.

Perthite and antiperthite. Potassium- and sodium-rich areas forming parallel or subparallel intergrowths are called perthite if the K-rich phase appears to be the host from which the Na-rich phase exsolved; an intergrowth is called antiperthite if the Na-rich phase appears to be the

host. Corresponding to the size of the exsolved areas a distinction is made between perthites (visible to the naked eye), microperthites (visible in the microscope), and cryptoperthites (detectable only by x-rays) In perthites and microperthites, the exsolved Na-rich phase usually occurs as albite forming somewhat irregular bands that are more or less parallel to a plane near ($\bar{8}01$). Within these bands the albite usually shows multiple twinning after the albite law. The K-rich host has usually changed into microcline (mostly into maximum microcline). The cryptoperthites can be rather complex. In these cases the K-rich host can be sanidine, orthoclase, or microcline; the Na-rich exsolved domains can be material with properties near those of analbite (they are then usually oriented after pericline-twin law) or near those of *albite* (then usually in albite-twin law orientation). The cryptoperthites frequently display a bluish to *whitish-milky* luster. If they are flawless and possess special qualities, they are also called moonstone and are used as gemstones (for example, [illegible]

ap [illegible]

$Or_{10}Ab_{90}$ ± An content, up to approximately 10%). In one way or another they show deviations from monoclinic *symmetry* As mentioned above, Or and Ab form a complete series of monoclinic solid solutions (sanidine-monalbite). *When cooled*, K-rich members change slowly by diffusive transformation into microcline. Na-rich members, however, change quickly into K-analbite by displacive transformation and in addition show a slow *diffusive* transformation tending to produce albite structure. The temperature of the displacive *transformation* monoclinic ↔ triclinic (monalbite ↔ analbite) is a function of the Ab/Or ratio and of the degree of Al/Si order. The larger these values, the *higher is* the temperature required. Exact data are not yet known. Whereas no displacive transformation has yet been found to occur *above room temperature* [illegible]

[illegible] temperature and Ab/(Ab + Or) ratio shown by natural material has been reported to be approximately as follows: Ab/(Ab + Or) ratio = 0.7, 0.8, 0.9, at 250, 500, 750C, respectively. During cooling both the displacive and the diffusive transformation processes can act concurrently in the formation of the final structural state of Na-rich alkali feldspars as found *in nature*. Therefore, the properties and state of twinning (visible lamellae or submicroscopical twinning) vary considerably from locality to locality. *The situation* is further complicated by the fact that under equilibrium conditions alkali feldspar of originally intermediate composition may have been exsolved into potassium and sodium feldspar, each of relatively pure composition (see *discussion of perthite* and antiperthite above). Because the relative speed of the three processes—displacive transformation, diffusive transformation, and exsolution—depends on several factors such as bulk chemical composition, rate of cooling, and others that are not yet known, the optical and structural properties of the anorthoclases vary strongly from crystal to crystal and can even within the same crystal be different from point to point. It must be emphasized, therefore, that anorthoclase is not a name for a particular phase with definite properties but usually applies to a mixture of several phases which may not even have a stability field of their own at any temperature.

PLAGIOCLASE FELDSPARS

Plagioclase feldspars form at high temperature a complete series of solid solutions, ranging from $NaAlSi_3O_8$(Ab) to $CaAl_2Si_2O_8$(An). The Ab end may take up to approximately 10 mole % $KAlSi_3O_8$(Or) into solid solution. Under equilibrium conditions they have a rather disordered Al/Si distribution near their melting points (~ 1100 to 1544°C). At lower temperatures the distributions become more ordered. This is the reason for the existence of so-called high-temperature and low-temperature optics connected by intermediate optics. The differing optics met with in natural materials can also be produced by heat treatment of material that has low-temperature optics. Besides the Al/Si distribution the chemical composition has an essential influence on the optics This fact has played a very important role in the development of microscopical methods for determining the chemical composition of plagioclases

The plagioclase series is arbitrarily subdivided and named according to increasing An content as follows: $Ab_{100}An_0$–$Ab_{90}An_{10}$ (albite); $Ab_{90}An_{10}$–$Ab_{70}An_{30}$ (oligoclase); $Ab_{70}An_{30}$–$Ab_{50}An_{50}$ (andesine); $Ab_{50}An_{50}$–$Ab_{30}An_{70}$ (labradorite); $Ab_{30}An_{70}$–$Ab_{10}An_{90}$ (bytownite); and $Ab_{10}An_{90}$–Ab_0An_{100} (anorthite). Because the content of silicic acid (in its anhydrous form SiO_2) is relatively higher in $NaAlSi_3O_8$ than in $CaAl_2Si_2O_8$, the Ab-rich members albite-oligoclase are frequently called acid plagioclases and the An-rich members bytownite-anorthite basic plagioclases. It is important to note that in contrast to the alkali feldspars in which the Al/Si ratio is always 1:3, this ratio varies in the plagioclases with increasing An content from 1:3 to 1:1. In consequence, solid state reactions involving homogenization or unmixing as a function of temperature take place much more sluggishly than in the alkali feldspars. Whereas in the latter a compositional change involves only a Na/K exchange, a Na/Ca exchange in the plagioclases necessitates a change of the Al/Si ratio as well. The Al and Si atoms have to cover quite considerable distances by diffusion and require correspondingly long times for such processes. It is therefore not astonishing to find most natural plagioclases zoned, that is, often to find great changes in the Ab/An ratio between the core and the rim of crystals. So-called normal zoning obtains if the core is richer in An than the rim; reverse zoning if the opposite is the case. Another consequence is the absence in nature of exsolution structures

on a visible scale between intergrown Ab- and An-rich parts, resembling those of the perthites. There are, however, many signs indicating that plagioclases of intermediate Ab/An ratio are not stable at low temperature, and indeed, submicroscopic exsolutions have been detected by the application of x-rays as discussed below.

In the sections on the alkali feldspars an account was given of the structural variations which are caused by the various possibilities of Al/Si distribution. The same considerations hold true in principle for the plagioclases. With these the conditions are, however, even more complex to survey as the varying Al/Si ratio now introduces a new factor. For this reason the phase relations within the plagioclase series are at present less well known than is the case of the alkali feldspars. This discussion is limited to listing the main structural varieties.

1. Phases with An content in states (*a*) similar to those of albite (low, intermediate, high); (*b*) similar to those of analbite; and (*c*) similar to those of monalbite.
2. Labradorite state (see below).
3. Anorthite state low.
4. Anorthite state high.

The states (1*a*), (1*b*), and (1*c*) have been discussed above (albite, analbite, monalbite). How far they extend into the plagioclase region under equilibrium conditions while retaining their main features is still unknown. In all cases, however, the increase of the Al/Si ratio with increasing An content adds additional possibilities of order/disorder. There are indications that the states (1*b*) and (1*c*) may exist up to approximately 20% An content. State (1*a*) may extend as calcium albite (intermediate to high) at intermediate and high temperatures up to approximately 80% An. State (2) appears to be a preliminary state of unmixing on a infrasubmicroscopic scale which is responsible for the labrador luster and develops from state (1*a*) during the course of cooling. States (3) and (4) are characterized by a new type of Al/Si ordering, caused by the Al/Si ratio approaching the value 1:1 which is reached exactly in pure $CaAl_2Si_2O_8$ (anorthite). X-ray photographs of states (3) and (4) show superstructure reflections, the character of which (necessitating a doubling of the *c* axis) can be explained by the experimentally supported assumption that each SiO_4 tetrahedron is surrounded by four AlO_4 tetrahedrons and vice versa. States (3) and (4) are distinguished by positional differences of the Ca ions within the voids of the framework. In pure $CaAl_2Si_2O_8$ state (3) is stable up to approximately 1000°C and has fixed Ca positions. Above this temperature, state (3) changes continuously and reversibly into state (4) in which the Ca ions occupy several positions more or less randomly. The transformation temperature at which state (3) inverts to state (4) decreases rapidly with increasing Ab content of the anorthite.

Albite. The name usually given by petrographers and mineralogists to plagioclases ranging from $Ab_{100}An_0$ to $Ab_{90}An_{10}$ is albite. The structure of albite and its high-temperature polymorphs was discussed above. Whereas the pure $NaAlSi_3O_8$ can be ideally ordered as case (3)tr of the table, any An content (as $CaAl_2Si_2O_8$) necessarily introduces Al/Si disorder. Therefore the properties of the low- and high-temperature modifications approach each other with increasing An content. Natural albite with more than 2 mole % An is usually unmixed on a submicroscopic scale to be discussed below. In addition to the remarks on albite above, some varieties may also be discussed here

Cleavelandite. This is structurally albite (low) and almost pure $NaAlSi_3O_8$. It has a tabular habit and is twinned after the albite law. The tabular individuals frequently show mosaic developments and present a bent appearance. They tend to occur as fan-shaped aggregates formed during the late pegmatitic stages of granite rock formation.

Peristerite. Natural plagioclases with an An content lying between $Ab_{98}An_2$ and $Ab_{85}An_{15}$ are usually submicroscopically unmixed into relatively pure albite Ab_mAn_n with $n < 2$ when $m + n = 100$, and into an oligoclase of the composition Ab_mAn_n where $n = 25$–30. This unmixing is easily detected by x-rays. If this unmixing is accompanied by a blue luster, such albite-oligoclase material is called peristerite and used as a gemstone, falsely called moonstone if of attractive appearance.

Adventurine or sunstone. This is an albite or oligoclase or andesine with a reddish luster caused by thin visible flakes oriented parallel to several structurally defined planes. They are probably Fe_2O_3 formed by exsolution

Oligoclase. Oligoclase has the composition $Ab_{90}An_{10}$ to $Ab_{70}An_{30}$. For its phase relations to the other plagioclases see the introductory discussion of plagioclase feldspars; for its unmixing exsolution processes, see the discussion of peristerite and sunstone.

Andesine. Andesine has the composition $Ab_{70}An_{30}$ to $Ab_{50}An_{50}$. For its phase relations to the other plagioclases see the discussion of plagioclase feldspars; exsolution and unmixing processes are discussed in the sections on sunstone and labradorite.

Labradorite. Labradorite has the composition $Ab_{50}An_{50}$ to $Ab_{30}An_{70}$. Its phase relations to the other plagioclases have been covered in the introductory discussion of plagioclase feldspars. Some additional remarks may be made as follows: x-ray photographs of albite differ from those of anorthite in that more reflections are present in the case of anorthite. In the literature these additional reflections are called *b* reflections (caused by Al/Si distribution) and *c* reflections (caused by the Ca position). These additional reflections indicate that the anorthite structure is a superstructure of the albite type. An explanation for this was given above. X-ray photographs of natural labradorite normally show *b*-type reflections also. However, in the range of approximately $Ab_{60}An_{20}$ up to $Ab_{25}An_{75}$ these *b* reflections are split into two, the degree of splitting increasing with increasing Ab

content. Parallel to this increase of splitting goes a decrease of the relative intensities and an increase of the diffuseness of the split *b* reflections. This behavior indicates that labradorites and the contiguous plagioclases (ranging in composition from approximately $Ab_{80}An_{20}$ up to $Ab_{25}An_{75}$) are not stable as solid solutions at low temperature. It appears that the labradorite state, state (2), is the beginning of an unmixing on a submicroscopical scale (in terms of angstrom units, A) into an oligoclase and a bytownite. Whether such a combination of oligoclase and bytownite is stable at low temperature under equilibrium conditions is not yet known. Obviously, such a mixture appears to be more stable at low temperature than the solid solution. But it might nevertheless be unstable with respect to other non-all-feldspar combinations within the quaternary system Na_2O-CaO-Al_2O_3-SiO_2. Numerous observations of natural alterations of the chemically intermediate plagioclases into products (saussurite) composed of zoisite-epidote, scapolite, and albite or into combinations of epidote, sericite, albite, calcite, and other minerals can be interpreted as indications confirming such a point of view, even when it is admitted that the presence of water appears to be essential for the alterations mentioned.

Bytownite. This mineral has the composition $Ab_{30}An_{70}$ to $Ab_{10}An_{90}$. For its phase relations to the other plagioclases and for its structure see the general remarks under plagioclase feldspar and the discussion of labradorite above.

Anorthite. Anorthite is the most basic member of the plagioclases, its composition ranging from $Ab_{10}An_{90}$ to pure An ($CaAl_2Si_2O_8$). Its phase relations to the other plagioclases and its structure are discussed in the sections on plagioclase feldspar and labradorite. It occurs in basic rocks (gabbro, norite, anorthosite); rarely as well-developed druse mineral (for example, in ejectamenta of the Somma, Vesuvius); sometimes in tuffs (good crystals are known from Miyaka, Japan); and very rarely in metamorphic rocks. [F.LA.]

Bibliography: W. H. J. Eitel, *Structural conversions in crystalline systems and their importance for geological problems*, Geol. Soc. Am. Spec. Paper 66, 1958; E. W. Heinrich, *Microscopic Petrography*, 1956; E. H. Kraus, W. F. Hunt, and L. S. Ramsdell, *Mineralogy*, 5th ed., 1959; A. N. Winchell and H. Winchell, *Elements of Optical Mineralogy*, pt. 2, 4th ed., 1951; for further information, readers are referred to original papers published since 1950 by S. W. Bailey, N. L. Bowen, M. G. Bown, W. L. Brown, R. B. Ferguson, P. Gay, J. R. Goldsmith, F. Laves, W. S. Mackenzie, I. D. Muir, T. R. Schneider, J. V. Smith, W. H. Taylor, R. J. Traill, and O. F. Tuttle in the following journals: *Acta Cryst.*, *Am. J. Sci.*, *Am. Mineral.*, *Geochim. et Coschim. Acta*, *J. Geol.*, *Mineral. Mag.*, and *Zeitschr. Kristallogr.*

Feldspathoid

A member of the feldspathoid group of minerals. The feldspathoids, like the feldspars, are aluminum silicates of sodium, potassium, and calcium, but contain a lower percentage of silica, reflecting the deficiency of silica in the magmas from which they crystallized. They may be found in the same rocks with feldspars but never with quartz. Unlike the feldspars, the feldspathoids are relatively rare but are of considerable importance in the study of the genesis of igneous rocks. See IGNEOUS ROCKS; SILICATE MINERALS.

The members of the feldspathoid group include nepheline, sodalite, cancrinite, leucite, lazurite, hauynite, and noselite. Nepheline is the most abundant and in places, as on the Kola Peninsula, U.S.S.R., is a principal constituent of large masses of nepheline syenite. Nepheline phonolites, the fine-grained equivalent of nepheline syenite, is less common. Sodalite, cancrinite, hauynite, and noselite are associated with nepheline in nepheline syenites. Leucite is confined to lavas rich in potassium and low in silica, the major occurrence of which is in central Italy. Lazurite, unlike the other members of the group, occurs characteristically in crystalline limestones. See CANCRINITE; LAZURITE; LEUCITE; see also LEUCITE ROCK; NEPHELINE SYENITE; NEPHELINITE. [C.S.HU.]

Felsite

An igneous rock with a felsitic or aphanitic (not visibly crystalline) texture, composed largely of light-colored (felsic) minerals (quartz and feldspar). The constituents may consist of glass, very fine crystalline material, or both. If relatively large crystals (generally visible) are abundant in the aphanitic portion, the rock becomes a felsite porphyry. Felsite is an aphanite in which felsic minerals predominate. See APHANITE; IGNEOUS ROCKS; PORPHYRY. [C.A.CA.]

Fennel

A culinary spice which has a long and interesting history. The plant, *Foeniculum vulgare*, a tall perennial herb, is a native of the Mediterranean re-



Fennel, *Foeniculum vulgare*. (USDA)

gion. It now occurs in all parts of the world, often as an escape from cultivation. The fruits are used in cookery, confectionery, and for flavoring beverages. In modern French and Italian cooking, fennel is indispensable. Oil of fennel is used in medicine, soaps, and perfumes. *See* SPICE AND FLAVORING; UMBELLALES. [P.D.S.]

Ferberite

A mineral with chemical composition $FeWO_4$. Ferberite is the iron member of the wolframite solid-solution series. It generally contains some manganese. It occurs in black, monoclinic, short, striated, prismatic crystals. Fracture is uneven. Luster is submetallic. Hardness is 4.5 on Mohs scale and specific gravity is 7.5. Streak is brownish black. It fuses readily to a magnetic globule. It differs from the manganese member huebnerite in that it is slightly harder, heavier, darker in color, and nearly opaque. For occurrence, test for tungsten, and use, *see* WOLFRAMITE. [E.C.T.C.]

Fer-de-lance

A large, slender, lance-headed viper, *Bothrops atrox*, found in the West Indies, southern Mexico, and tropical South America. It is called the barba amarilla in Central America and the jararaca in Brazil. This is probably the most dangerous of the Central American snakes, both because of its long fangs and size (up to 8 ft) and because of the character of its venom. It is nocturnal and spends much of its time in the burrows of other animals. There are about 36 species of the genus *Bothrops*, widely distributed in tropical America.

The fer-de-lance is highly variable in color, ranging from gray through olive or brown to reddish, and marked with dark cross bands It produces large litters of living young; a big female may give birth to 60 or 70 young at a time. *See* SQUAMATA. [J.D.B.]

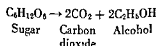
Fermentation

A term used since 1600 to denote a decomposition of foodstuffs generally accompanied by the evolution of gas. The best-known example is alcoholic fermentation in which sugar is converted into alcohol and carbon dioxide. This conversion, described by the equation

$$\underset{\text{Sugar}}{C_6H_{12}O_6} \rightarrow \underset{\text{Carbon dioxide}}{2CO_2} + \underset{\text{Alcohol}}{2C_2H_5OH}$$

was established by J. Gay-Lussac in 1815.

Before 1800 the association of yeast or leaven with fermentation had been noted; but the nature of these agents was not understood. Experiments of C. Cagniard-Latour, of F. T. Kützing, and of T. Schwann in 1837 indicated that yeast is a living organism and is the cause of fermentation.



This view was opposed by such leading chemists as J. von Liebig and F. Wöhler, who sought a chemical rather than a biological explanation of the process. The biological concept became generally accepted following the work of Louis Pasteur, who concluded that fermentation is a physiological [illegible]

parable processes; both represent decompositions of organic matter brought about by microorganisms in the absence of air. The difference is determined by the nature of the decomposable material; sugary substances generally yield products with pleasant odor and taste (fermentation), whereas proteins give rise to evil-smelling products (putrefaction).

Pasteur also discovered the lactic acid and butyric acid fermentations, and from his experiments concluded that each kind of fermentation was caused by a specific microbe. Later work has supported this idea to a large extent, and considerably increased the number of specific fermentations. *See* LACTIC ACID.

During fermentation, organic matter is decomposed in the absence of air (oxygen); hence, there is always an accumulation of reduction products, or incomplete oxidation products. Some of these products (for example, alcohol and lactic acid) are of importance to man, and fermentation has therefore been used for their manufacture on an industrial scale. There are also many microbiological processes that go on in the presence of air while yielding incomplete oxidation products. Good examples are the formation of acetic acid (vinegar) from alcohol by vinegar bacteria, and of citric acid from sugar by certain molds (for example, *Aspergillis niger*). These microbial processes, too, have gained industrial importance, and are often referred to as fermentations, even though they do not conform to Pasteur's concept of fermentation as a decomposition in the absence of air. For the mechanism of fermentation *see* BACTERIAL METABOLISM. For details of various fermentation processes *see* INDUSTRIAL MICROBIOLOGY. [C.B.V.N.]

Bibliography: R. Dubos, *Louis Pasteur, Free Lance of Science*, 1950; J. W. Foster, *Chemical Activities of Fungi*, 1949; S. C. Prescott and C. G. Dunn, *Industrial Microbiology*, 3d ed., 1959.

Fermi-Dirac statistics

The statistical description of particles or systems of particles that satisfy the Pauli exclusion principle (*see* EXCLUSION PRINCIPLE). This description was first given by E. Fermi, who applied the Pauli exclusion principle to the translational energy levels of a system of electrons. It was later shown by P. A. M. Dirac that this form of statistics is also obtained when the total wave function of the system is antisymmetrical. *See* QUANTUM THEORY, NONRELATIVISTIC.

Distribution function. Such a system is described by a set of occupation numbers $\{n_i\}$ which specify the number of particles in energy levels ϵ_i. It is important to keep in mind that ϵ_i represents a finite range of energies, which in general contains a number, say g_i, of nondegenerate quantum states. In the Fermi statistics, at most one particle is allowed in a nondegenerate state. (If spin is [illegible] into account, two particles may be contair[illegible]

such a state.) This is simply a restatement of the Pauli exclusion principle, and means that $n_i \leqq g_i$. The probability of having a set $\{n_i\}$ distributed over the levels ϵ_i, which contain g_i nondegenerate levels, is

$$W = \prod_i \frac{g_i!}{(g_i - n_i)!n_i!} \tag{1}$$

(*see* Boltzmann statistics). Equation (1) gives just the number of ways in which one can pick n_i out of g_i, which is intuitively what one expects for such a probability. In Boltzmann statistics this same probability is given by

$$W' = \prod_i \frac{g_i^{n_i}}{n_i!} \tag{2}$$

The equilibrium state which actually exists is the set of ns that makes Eq. (1) a maximum, under the auxiliary conditions:

$$\Sigma n_i = N \tag{3a}$$

$$\Sigma n_i \epsilon_i = E \tag{3b}$$

These conditions express the fact that the total energy E and the total number of particles N are given.

One finds for this most probable distribution:

$$n_i = \frac{g_i}{\frac{1}{A} e^{\beta \epsilon_i} + 1} \tag{4}$$

Here A and β are parameters, to be determined from Eq (3); in fact, $\beta = 1/kT$ where k is Boltzmann's constant and T is the absolute temperature. When the 1 in the numerator may be neglected, Eq. (4) goes over into the Boltzmann distribution; this provides a procedure for identifying β. It is known that in classical statistics the Boltzmann distribution may be obtained if specific assumptions are made as to the number of collisions taking place. It is there assumed that the number of collisions per second in which molecules with velocities in cells i and j in phase space produce molecules with velocities in cells k and l is given by

$$A_{ij}^{kl} = n_i n_j a_{ij}^{kl} \tag{5}$$

Here A_{ij}^{kl} is a geometrical factor. This leads to the Boltzmann distribution. The Fermi distribution, Eq. (4), may be obtained if instead of Eq. (5) one assumes for the number of collisions:

$$A_{ij}^{kl} = a_{ij}^{kl} n_i n_j \left(\frac{g_k - n_k}{g_k}\right)\left(\frac{g_l - n_l}{g_l}\right) \tag{6}$$

One observes the interesting quantum theoretical feature that the probability for a collision depends on the occupation numbers of the states into which the colliding particles will go. In particular, if these final states are filled up ($n_k = g_k$), no collision with that state as a final state can occur.

The distribution f_i is often used, it is defined by

$$n_i = f_i g_i \tag{7}$$

Applications. For a system of N electrons, each of mass m in a volume V, Eq. (4) may be written as

$$f(v_x v_y v_z)\, dv_x\, dv_y\, dv_z = 2 \frac{m^3}{h} V \frac{dv_x\, dv_y\, dv_z}{\frac{1}{A} e^{mv^2/2kT} + 1} \tag{8}$$

where h is Planck's constant. Equations (3a) and (3b) may now be transformed into integrals, yielding

$$\frac{N}{V}\frac{\lambda^3}{2} = U_{1/2}(A) \tag{9a}$$

$$\frac{E}{3kT}\frac{\lambda^3}{V} = U_{3/2}(A) \tag{9b}$$

Here $\lambda = (h^2/2\pi mkT)^{1/2}$ is the thermal de Broglie wavelength, and $U_\rho(A)$ is the Sommerfeld integral defined by

$$U_\rho(A) = \frac{1}{\Gamma(\rho + 1)} \int_0^\infty \frac{u^\rho\, du}{\frac{1}{A} e^u + 1} \quad \text{where } u = \frac{1}{2}\frac{mv^2}{kT} \tag{10}$$

$$U_\rho(A) \cong A \quad (A \ll 1) \tag{10a}$$

$$U_\rho(A) \cong \frac{(\ln A)^{\rho+1}}{\Gamma(\rho + 1)} \tag{10b}$$

Here Γ is the usual Γ-function. Very often one writes instead of A in the Fermi distribution a quantity μ defined by

$$\mu = kT \ln A \qquad A = e^{\mu/kT} \tag{11}$$

It may be shown (for instance, by going to the classical limit) that μ is the chemical potential. It may be seen that if ϵ is large, $1/kT(\epsilon - \mu) \gg 1$; hence the Fermi distribution goes over into a Maxwell Boltzmann distribution. It is easy to verify that this inequality may be transcribed so as to state

$$\frac{V}{N} \gg \lambda^3 \tag{12}$$

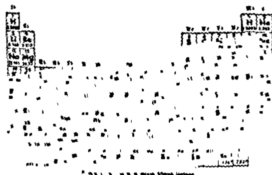
Physically this is reasonable because (12) says that classical conditions pertain when the volume per particle is much larger than the volume associated with the de Broglie wavelength of a particle. For example, $V/N\lambda^3$ is about 7.5 for helium gas (He^3) at 4°K and 1 atmosphere. This indicates that classical statistics may perhaps be applied, although quantum effects surely play a role. For electrons in a metal at 300°K, $V/N\lambda^3$ has the value 10^{-4}, showing that classical statistics fail altogether for electrons in metals. When the classical distribution fails, a degenerate Fermi distribution results. Numerically, if $A \gg 1$, a degenerate Fermi distribution results; if $A \ll 1$, the classical results are again obtained. For example, Eqs. (9a), (9b), and (10a) show that $E = 3/2NkT$ and the specific heat should be $3/2\ R = 3/2\ Nk$. However for an electron gas this does not apply since such a system is degenerate for normal temperatures and $A \gg 1$. A somewhat lengthy calculation yields the result that in that case ($A \gg 1$) the contribution

to the specific heat is negligible. This resolves an old paradox, for according to the classical equipartition law, the electronic specific heat C should be $3/2\ Nk$, whereas in reality it is very small; in fact, $C = \gamma T$, where γ is a very small constant. This is a consequence of the fact that an electron gas at normal temperature is a degenerate Fermi gas. The electrical resistance of a metal can be understood on a classical picture, but the lack of a specific heat is a pure quantum effect. *See* BOSE-EINSTEIN STATISTICS; KINETIC THEORY OF MATTER; QUANTUM STATISTICS. [M.DR.]

Bibliography: *See* STATISTICAL MECHANICS.

Fermium

Element number 100, fermium, Fm, is an element belonging to the actinide series. Fermium does not occur in nature; its discovery and production have been accomplished by artificial nuclear transmutation of lighter elements. Isotopes of mass number



2[illegible] have been discovered; all are radioactive, decaying with half-lives ranging from less than 1 min to a few days. The total weight of fermium which has been synthesized is much less than one-millionth of a gram. *See* ACTINIDE ELEMENTS; RADIOACTIVITY.

The first *isotope of fermium to be discovered* was Fm^{255}. In 1952, scientists of the Los Alamos, Argonne, and University of California laboratories of the U.S. Atomic Energy Commission isolated Fm^{255} from the debris of a hydrogen-bomb explosion. Uranium present in the weapon underwent multiple neutron capture to give U^{255}; a subsequent chain of β-decays yielded finally the mass 255 isotope of fermium. Other fermium isotopes have been produced by bombardment of heavy element isotopes with helium, beryllium, carbon, and oxygen ions accelerated in cyclotrons or linear accelerators. For production of larger amounts of fermium isotopes, gram quantities of Pu^{239} or heavier isotopes are irradiated with neutrons for periods of several years in nuclear reactors. Isotopes of fermium and the other higher actinide elements are isolated as products of successive neutron capture and β-decay. *See* BERKELIUM.

Spontaneous fission with the emission of neutrons, which becomes increasingly important in the heavier elements, is the major mode of decay for Fm^{256}. Fermium-254 also decays to a small extent by spontaneous fission; its primary mode of decay is by emission of an α-particle.

Preliminary tracer-scale studies indicate that the chemical properties of fermium are those characteristic of the actinide elements. Fermium appears to exist only in the 3+ oxidation state in aqueous solution. Ion-exchange chromatography has played a significant part in the separation and identification of fermium. Fermium is separated from other actinide elements when a complexing agent is passed through a column of cation-exchange resin onto which the actinides have been adsorbed. The elements are desorbed from the resin at different rates and leave the column in distinct fractions. *See* ION EXCHANGE; NUCLEAR CHEMISTRY; NUCLEAR REACTION; TRANSURANIUM ELEMENTS. [S.G.T.]

Bibliography: J. J. Katz and G. T. Seaborg, *The Chemistry of the Actinide Elements*, 1958.

Ferret

A carnivore of the family Mustelidae, closely related to the weasels. The common ferret, *Mustela fure*, is a native of Africa. It is now rare in the wild state but is kept as a domestic animal in America and Europe. It is an excellent rat killer and has been employed as a hunter, especially for *rabbits and other burrowing animals, being able to* pursue them underground because of its long, slender body and short legs.



The domestic ferret, *Mustela fure*; length 12–14 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

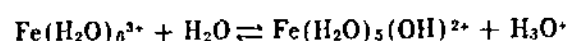
The black-footed ferret, *Mustela nigripes*, almost extinct, is a large, brown weasel formerly common as a predator in the prairie-dog towns of the West. *See* CARNIVORA; PRAIRIE DOG; WEASEL. [J.D.B.]

Ferric compound

A compound containing iron in the 3+ oxidation state. The ferric, iron(III), state is obtained when iron is dissolved in oxidizing acids or when ferrous solutions are oxidized.

Although ferrous hydroxide is a fairly strong base, ferric hydroxide is much weaker. In fact the

hydrated ion acts as an acid in the following manner.



$$Fe(H_2O)_6^{3+} + H_2O \rightleftharpoons Fe(H_2O)_5(OH)^{2+} + H_3O^+$$

Ferric salts are easily hydrolyzed in this manner to produce, after two more steps similar to the above equation, $Fe(H_2O)_3(OH)_3$ or $Fe(OH)_3 \cdot 3H_2O$.

Because the composition of this red-brown gelatinous precipitate varies, it also is called a hydrous ferric oxide and written $Fe_2O_3 \cdot xH_2O$. The presence of this hydrous oxide in colloidal form gives ferric solutions their characteristic yellow and often muddy appearance.

The oxide, Fe_2O_3, known under many descriptive names, is used as a paint pigment, polishing compound, and rouge.

Ferric chloride, $FeCl_3$, is covalent in nature when anhydrous and gas-density measurements indicate that the formula of the gas is Fe_2Cl_6. It is used as a coagulant and as a mordant and etching compound. A double salt, $(NH_4)_2SO_4 \cdot Fe_2(SO_4)_3 \cdot 24H_2O$, is one of a series of compounds called alums. *See* FERROUS COMPOUND; IRON. [E.E.WR.]

Ferricyanide

A compound containing the complex ion, $[Fe(CN)_6]^{3-}$, where iron has an oxidation state of 3+. The coordination [illegible]

Using the nomenclature of coordination compounds, $Na_3[Fe(CN)_6]$ for example, should be called sodium hexacyanoferrate(III)

Although the ferricyanides of the alkali and alkaline-earth metals are water-soluble, the remainder are water-insoluble

Ferricyanides are prepared from ferrocyanides by oxidation of the iron with chlorine. The uses are much the same as for ferrocyanides. When a solution of ferrous salt is mixed with a solution of an alkali ferricyanide, a precipitate is obtained which is called Turnbull's blue and is not $Fe_3[Fe(CN)_6]_2$. It is thought to be the same as Prussian blue and may be represented as $K[Fe^{II}(CN)_6Fe^{III}]$. *See* FERROCYANIDE; IRON. [E.E.WR.]

Ferrimagnetic garnets

A class of ferrimagnetic oxide materials that have the same crystal structure as garnet. The crystal structure affords three types of nonequivalent sites in the oxygen lattice which are occupied by metal ions. The sites are referred to as the $16(a)$, $24(c)$, and $24(d)$ sites. A metal ion in a $16(a)$ site is surrounded by 6 oxygen ions which are at the corners of an octahedron. The metal ions in the $24(d)$ sites are found in nonregular tetrahedrons formed by oxygen ions, while the metal ions in the $24(c)$ sites have 8 oxygen ions as nearest neighbors which can be thought of as being at the corners of a distorted cube. *See* CRYSTALLOGRAPHY; FERRIMAGNETISM; FERRITE.

The ferrimagnetic garnets are of great theoretical interest because they have a highly ordered structure and because they accommodate the large rare-earth ions, some of which have a small contribution to their magnetism due to the orbital motion of electrons in addition to the magnetism due to electron spin. The practical engineering interest in the ferrimagnetic garnets is due almost entirely to one member of the family, yttrium iron garnet. This material has important engineering advantages for certain microwave ferrite devices because it has the narrowest ferromagnetic resonance absorption line ever measured (less than one oersted at 10,000 Mc). *See* FERRITE DEVICES; GYRATOR.

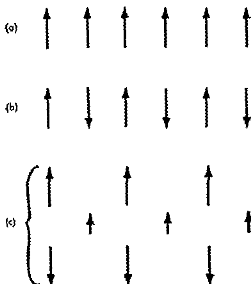
The typical formula of a rare-earth iron garnet is $3X_2O_3 \cdot 5Fe_2O_3$ where the trivalent X ion is yttrium or any of the rare-earth ions with an atomic number above 61. (The rare-earth ions below samarium apparently have ionic radii too large to fit into the garnet structure). The iron occupies both the $16(a)$ and $24(d)$ sites while the rare-earth or yttrium ion occupies the $24(c)$ sites. The 24 iron ions (per unit cell) on the (d) sites are strongly coupled, with an antiferromagnetic coupling, to the 16 iron ions in the (a) sites. If the 24 ions on the (c) sites have a magnetic moment, it is weakly aligned antiparallel to the moment of the ions on the (d) sites. In the case of yttrium iron garnet, the yttrium ion is diamagnetic; so the magnetism is due to the trivalent iron ions alone. [C.L.H.]

Bibliography: M. L. Keith and R. Roy, Structural relations among double oxides of trivalent elements, *Am. Mineralogist*, 39(1–2):1–23, 1954; J. W. Nielsen and E. F. Dearborn, Growth of single crystals of magnetic garnets, *J. Phys. Chem. Solids*, 5(3):202–207, 1958; L. G. Van Uitert, F. W. Swanekamp, and F. R. Monforte, Method for forming large ferrite parts for microwave applications, *J. Appl. Phy.*, 27(11):1385–1387, 1956; L. G. Van Uitert and F. W. Swanekamp, Resistivity and density of polycrystalline yttrium iron garnet, *J. Appl. Phy.*, 28(12):1513–1514, 1957; W. P. Wolf and G. P. Rodrigue, Preparation of polycrystalline ferrimagnetic garnets materials for microwave applications, *J. Appl. Phy.*, 29(1):105–108, 1958; H. S. Yoder and M. L. Keith, Complete substitution of aluminum for silicon, *Am. Mineralogist*, 36:519, 1951.

Ferrimagnetism

A type of permanent magnetism in which the magnetic moments of neighboring ions have a tendency to align antiparallel. It occurs principally in the compounds known as ferrites (*see* FERRITE). The term ferrimagnetism was coined by L. Néel in 1948.

As shown in part *c* of the figure, in ferrimagnetism the atomic magnetic moments, at sufficiently low temperatures (below the so-called Curie temperature), point up and down in an ordered array so as to cancel partially. The net uncanceled moments provide the macroscopic crystal with mag-



Schematic representation of arrangements of atomic magnetic moments in (a) ferromagnetism, (b) antiferromagnetism, (c) ferrimagnetism.

netic properties very similar to those of ferromagnetism. See FERROMAGNETISM.

The simplest magnetic variety of ferrite can be represented by the chemical formula $MOFe_2O_3$, where M is a divalent metal ion. The crystal structure (spinel) has two types of sites for the metal ions: (1) tetrahedral sites, in which each metal ion is surrounded by four oxygen ions; and (2) octahedral sites, in which each metal ion is surrounded by six oxygen ions. All ferrimagnetic materials contain more than one kind of magnetic ion, or more than one kind of crystallographic site, or both. This introduces an asymmetry into the elementary interactions which is responsible for the peculiar arrangement of atomic moments.

In Néel's theory of ferrimagnetism, the magnetic structure of magnetite, for example, is assumed to be $Fe^{+++}(\uparrow)Fe^{++}(\uparrow)Fe^{+++}(\downarrow)O_4$, where the arrows represent directions of magnetic moments of the metal ions, as in the figure. The Fe^{++} ion has spin 2 and magnetic moment of 4 Bohr magnetons; the Fe^{+++} ions have spin 5/2 and magnetic moment of 5 Bohr magnetons. Thus, the total magnetic moment per molecule of magnetite should be $5 + 4 - 5 = 4$ Bohr magnetons; the experimental value of 4.1 is in close agreement. The theory has also been confirmed directly by neutron diffraction (see NEUTRON DIFFRACTION).

The parallel and antiparallel couplings of the magnetic moments are probably produced by superexchange through the oxygen ions, as in antiferromagnetism.

Néel divides the ferrite into magnetic sublattices and assumes a Weiss-type molecular field interaction between the sublattices. The theory is very similar to the molecular-field theory of antiferromagnetism and accounts very well for the gross magnetic behavior of ferrites. See ANTIFERROMAGNETISM. [E.A.; F.K.]

Bibliography: C. Kittel, *Introduction to Solid State Physics*, 2d ed., 1956; J. S. Smart, The Néel theory of ferrimagnetism, *Am. J. Phys.*, 23(6): 356–370, 1955.

Ferrite

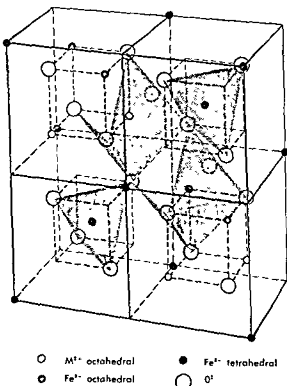
A chemical compound, often magnetic, of which ferric oxide (Fe_2O_3) is a component. The name is sometimes limited to those ferrites, such as $NiFe_2O_4$, which have the spinel structure. Many of *the magnetic ferrites are important in industry and* in the study of the magnetic forces between atoms in solids. See CRYSTAL STRUCTURE; FERRITE DEVICES; MAGNETIC MATERIALS.

The principal ferrites have the following chemical formulas and crystal structures (the symbol M indicates a divalent metal ion): (1) MFe_2O_4, spinel; (2) $MO \cdot 6Fe_2O_3$, magnetoplumbite; (3) $MFeO_3$ (orthoferrite), distorted perovskite; and (4) $M_3Fe_5O_{12}$, garnet. In (1), one often finds M = Mn, Fe, Co, Ni, Cu, Mg, or Zn; in (2), M = Ca, Ba, or Pb; in (3) and (4), M = Y, Gd, or some other rare earth; and in (4), Fe may be replaced by a variety of other metals.

Ferrites are usually prepared by heating mixtures of the component oxides or carbonates in an atmosphere of nitrogen, oxygen, or air at 1100–1400°C. They are often ground and pressed after the first firing and refired. Ferrites are mechanically hard and can be strong or brittle.

Superexchange forces. The magnetic ferrites are generally ferrimagnetic, the magnetic moments of some of the metal ions being held antiparallel to the others by the action of the quantum mechanical forces of superexchange (*see* ANTIFERROMAGNETISM; FERRIMAGNETISM). *Consider* $NiFe_2O_4$ *as an* example. This has a cubic crystalline structure (see illustration) such that half of the Fe^{3+} ions are surrounded by four O^{2-} ions (A positions); the other Fe^{3+} ions and all the Ni^{2+} ions are surrounded by six O^{2-} ions (B positions). This is better shown when the formula is written $(Fe)(NiFe)O_4$ to indicate the atoms that are in A and in B positions. The strong superexchange force holds the moments of the ions in the B positions antiparallel to those in the A positions so that the net moment of one molecule is $(5 + 2) - 5 = 2$ Bohr units (5 for Fe^{3+} and 2 for Ni^{2+}). Theory and experiment are in approximate agreement when M = Mn, Fe, Co, Ni, Cu, or Zn, the net moments per molecule being 5, 4, 3, 2, 1, and 0. More complicated and only partially understood situations occur when nonmagnetic ions such as Zn are added to MFe_2O_4 in various amounts; in these cases, the net moment of MFe_2O_4 is raised because the Zn^{2+} ions go into the A positions, some of the Fe^{3+} ions transfer to the B positions, and the imbalance of moments in the A and B positions thus increases.

The ferrites of the magnetoplumbite type are hexagonal in structure. Orthoferrites have ortho-



Structure of inverse spinel ferrite, MFe_2O_4 or (Fe) $(MFe)O_4$. One half of a unit cell is shown; the whole structure is a 3-dimensional checkerboard arrangement of the small cubes, of which 4 are drawn.

rhombic crystal structure and weak ferromagnetism at room temperature; this can be understood in terms of the superexchange as modified by the orthorhombic crystalline field. Many of the orthoferrites are strongly ferromagnetic at liquid helium temperatures. The garnets have the well known cubic structure of ordinary garnet, $Ca_3Fe_2Si_3O_{12}$.

Commercial types. The most used ferrites are the spinels. Commercial ferrites are of several kinds: (1) $Mn_{0.5}Zn_{0.5}Fe_2O_4$ and $Ni_{0.5}Zn_{0.5}Fe_2O_4$ (approximate compositions) have high permeability and low hysteresis loss and are used in high frequency transformers and filters. The electrical resistivities are about 10^2 and 10^5 ohm-cm, respectively (2) $Mg_{0.5}Mn_{0.5}Fe_2O_4$ has a rectangular hysteresis loop and is used for the storage of information. It has a resistivity of the order of 10^8 or 10^9 ohm-cm, and in high-speed devices the magnetic flux can be reversed within a time of the order of 1 microsecond. (3) $BaO \cdot 6Fe_2O_3$, in the form of a sintered fine powder, is a good permanent magnet material. It has high magnetic crystal anisotropy, and has the best quality when the crystals are aligned by pressing so that their axes are parallel to each other. $Co_{0.5}Fe_{2.5}O_4$, although not now made commercially to any considerable extent, was the first ferrite used as permanent magnet material. Commercial ferrites of the spinel type are often known as ferroxcubes; those of composition $BaO \cdot 6Fe_2O_3$ as ferroxdurs.

A number of ferrites of the spinel and garnet types are being used experimentally in gyrators and other microwave devices.

Lattice positions. When other elements are added as oxides to the basic types of ferrites, it is an important consideration to know what positions they occupy in the crystal lattice. The addition of zinc to the spinel ferrites has already been discussed. In ferrites of the garnet type, the positions occupied by the metal ions are of three kinds, designated A, B, and C, in which these ions are surrounded by tetrahedra, octahedra, and dodecahedra of O^{2-} ions, respectively. The position preferred by a given ion gives important information concerning the distribution of electrons about these ions. [R.M.BO.]

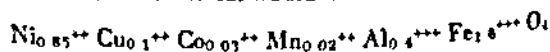
Bibliography: E. W. Gorter, *Philips Research Repts.*, 9:295, 403, 1954.

Ferrite devices

Electrical devices whose principle of operation is based upon the use of ferrites, which are ceramic-like materials with permeabilities ranging up to several thousand and with electrical resistivities up to 10^{11} times as high as metals. Thus it is not necessary to laminate or powder ferrite materials in order to reduce eddy-current losses, as is the case with metallic magnetic materials. *See* FERRIMAGNETISM; FERRITE; FERROMAGNETISM.

Spinel crystal structure. Although the first recorded observation by man of the ferromagnetic properties of matter were made by Thales of Miletus, circa 600 B.C., on naturally occurring ferrous ferrite (magnetite), modern technology had made little use of ferrites until the contributions of the Dutch scientists during World War II. The word ferrite as used here refers to those materials with the spinel crystal structure, magnesium aluminate, and with the chemical formula XFe_2O_4 where X represents any divalent metal ion whose size is such that it will fit into the crystal structure. In special cases, the divalent ion X could be replaced by an equal molar mixture of an univalent and trivalent ion. Thus, lithium ferrite can be thought of as having the formula $Li_{0.5}Fe_{0.5}Fe_2O_4$. However, to date the commercially practical ferrites are those in which the divalent X ion is either Mg, Mn, Fe, Co, Ni, Cu, Zn, Cd, or a mixture of these ions.

Some ferrites have, however, found important uses where some of the iron is replaced by some other trivalent ion such as aluminum. In fact, very complex ferrites are now manufactured for particular uses. As an example, nickel ferrite ($NiFe_2O_4$) has little practical value in certain microwave devices, but if this material is modified by adding cobalt to narrow the ferromagnetic resonance line width, adding aluminum to decrease the saturation moment, and adding copper and manganese to reduce dielectric loss, a valuable microwave ferrite material can be constructed. Following the above notation, this material would have the formula



$$Ni_{0.85}^{++}\,Cu_{0.1}^{++}\,Co_{0.03}^{++}\,Mn_{0.02}^{++}\,Al_{0.4}^{+++}\,Fe_{1.6}^{+++}\,O_4$$

Thus, in practice, complicated mixed crystals of the various possible ferrites are formed in order to produce a material that has a certain property maximized for a particular application, and quite a science has recently developed around the problem of tailoring the material to the application.

Examples of ferrite applications*

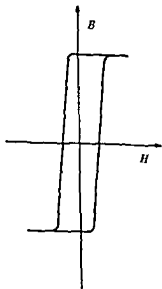
Type of component	Typical frequency	Characteristic sought	Favorable ferrite properties	Typical ferrite μ_0	Desired improvements
Linear *B-H*, low flux density					
Filter inductors	100 kc	High Q, precision, stability, low modulation	High μQ, low hysteresis, convenient core shapes	MnZn 1000–2500	Higher μQ, lower temperature coefficient
I-f transformers	465 kc	High Q, stability, adjustability	High μQ, shapes	NiZn† 100–200	Higher μQ to higher *frequencies*, better
Antenna cores	1000 kc	Moderate Q, stability	High μQ, rod shapes	NiZn† 100–200	stability to temperature and to magnetic and mechanical shock
Wide-band transformers	To 15 Mc	Reproducibility of transmission characteristics	High μ, low losses, assembled core structures	MnZn 1000–2500 NiZn† >500	Very uniform magnetic and mechanical properties
Adjustable inductors	Various	Adjustment ±20%	High μ, shapes	MnZn NiZn†	
Tuners	Various	Adjustment >10/1	Various mechanical structures, μ variation in dc field	MnZn NiZn	
Miniature inductors	Various	Moderate Q, small size	Cup shapes	MnZn >1000 NiZn† >1000	Higher saturation, less loss at moderate B_m
Loading coils	Voice	Stability, low modulations, low leakage flux	High μQ		Lower sensitivity to dc fields, lower hysteresis loss
Nonlinear *B-H*, medium to high flux densities					
Flyback transformers	15–100 kc	Low loss, moderate size		MnZn >750 NiZn† >750	Higher saturation, lower losses Higher Curie temperatures
Miniature transformers	Pulse	Small size, easy mounting	High μ, cup shapes	NiZn 4000	Higher saturation
Suppression beads		High resistive impedance	High loss above critical frequency	MnZn >500 NiZn† >500	
Recording heads			Mechanical rigidity, low loss	NiZn MnZn	Higher saturation, nonabrasive surface
Highly nonlinear, rectangular loop					
Memory cores	Pulse	Two identifiable stable states, fast flux reversal	Small rings practical, reasonable rectangularity	MgMn	Higher degree of rectangularity, stability, uniformity, low cost
Switching cores	Pulse	Fast flux reversal, *high output*	Reasonable rec*tangularity*	MgMn	High saturation for high output
Microwave properties					
Isolators, attenuators, circulators, switches, modulators			Faraday rotation, ferromagnetic resonance	MgMn†	
Magnetostriction and other properties					
Temperature controls		Sensitivity to temperature	Low Curie temperature, temperature sensitivity		

* Reproduced by permission of *Proceedings of IRE*.
† May contain additional elements such as Mg, Mn, Co, Cu, Al.

Applications. Ferrites have already found wide application because in certain applications they yield higher efficiency with smaller volume and lower costs. A summary of the applications of ferrites is given in the table.

In the design of inductors and transformers, the so-called μQ product of a material has been found to be a useful index of the quality of the material. In this product, μ is the permeability and Q is the figure of merit, a reciprocal measure of the core losses in the material (*see* Q). The factor Q is equal to $\omega L/R$, where R is the effective series resistance arising from core loss Obviously in inductor design, a material with a high μQ product is desirable. A high value of μ produces larger inductance per turn and thus allows a more efficient coil or a smaller coil. A high permeability also reduces leakage flux. The ferrite materials have the highest μQ product of any commercially available magnetic material today Typical μQ values measured at 20 kc for various materials are: iron dust, 2000; Permalloy powder, 4000, molybdenum Permalloy powder, 10,000, ½-mil molybdenum Permalloy tape, 100,000; and manganese zinc ferrite, 250,000.

Ferrite rods are widely used today for antenna cores in broadcast-radio receivers. Because of the high permeability of the ferrite rod at radio frequencies, the sensitivity of the antenna is greatly increased over the normal loop antenna with an air core. The antenna rods in use today are usually ⅛-¾ in. in diameter and 2-8 in long. The antenna winding is tightly wrapped over the length of the rod. In order to make use of the intrinsic permeability of the ferrite without deleterious demagnetizing effects, the ratio of the length to the diameter of the rod should be as large as possible.



Magnetic field density, B, as a function of magnetic field strength, H. This plot is called a hysteresis loop, the area encompassed by the loop being proportional to power loss within the ferrite.

The largest use of ferrite today is in flyback transformers for television picture tubes. The cores for these transformers must have low loss at high flux densities (up to 2000 gauss) and to frequencies as high as 100 kc, which is the effective flyback frequency used in scanning a television picture tube.

Ferrite computer elements. A rapidly growing and potentially large application for ferrites today is for logic and memory uses in modern digital computers. This is a natural use for ferrite cores, because they have two easily achieved stable states of magnetization. For instance, a toroidal core can be magnetized either clockwise or counterclockwise. The properties which are important for these applications are the rectangularity of the hysteresis loop (see illustration) and the coercive force of the material. The ideal core would have a perfectly rectangular B-H relationship and a coercive force so low that it could be easily driven from one state of magnetization to the other by high-speed transistors with only one turn on each core. *See* HYSTERESIS, MAGNETIC.

Recently, many proposals have been made for performing logic and memory functions using a single piece of ferrite with two or more apertures through which the magnetizing windings are threaded. These devices are easier to wire into a computer system and are capable of performing a wide variety of functions that are not easily accomplished by using individual cores. The most useful of these devices take the form of a ferrite plate with an array of holes in it, a "ladder" of ferrite with side rails and rungs and an ordinary core with a small secondary aperture in one side of the core.

Finally, the very low loss of certain ferrite materials has recently permitted the building of a new class of nonreciprocal microwave circuit elements which have found wide application. For a complete discussion of these devices, *see* GYRATOR. *See also* FERRIMAGNETIC GARNETS. [C.L.H.]

Bibliography: Ferrites issue, *Proc. IRE*, 44(10): 1233–1468, 1956; R. L. Harvey, I. J. Hegyi, and H. W. Leverenz, Ferromagnetic spinels for radio frequencies, *RCA Rev.* 11(3):321–363, 1950; C. L. Hogan, The microwave gyrator, *Bell System Tech. J.*, 31(1):1–31, 1952; J. A. Rajchman, A myriabit magnetic-core matrix memory, *Proc. IRE*, 41(10): 1407–1421, 1953; J. L. Snoek, *New Developments in Ferromagnetic Materials*, 1947.

Ferroalloy

An important group of metallic raw materials required for the steel industry. Ferroalloys are the principal source of such additions as silicon (Si) and manganese (Mn), which are required for even the simplest plain-carbon steels, and chromium (Cr), vanadium (V), tungsten (W), titanium (Ti), and molybdenum (Mo), which are used in both low- and high-alloy steels. Also included are many other more complex alloys, such as those for high-temperature use. Ferroalloys are unique in that they are brittle and otherwise unsuited for any

Analysis of typical ferroalloys, % weight

Type of ferro-alloy*	Mn	Si	C	Cr	Mo	Al	Ti	V
Ferromanganese								
Standard	78–82	1.25	7 5					
Med. carbon	80–85	1.25–2.5	1–3†					
Low carbon	80–85	1.25–7.0	0.75					
Ferrosilicon								
50% Regular		47–52	0.15					
75% Regular		73–78	0.15					
Ferrochromium								
High carbon		1–2	1.5–6 0†	67–70				
Low carbon		0 3–1.0	0.03–2.0†	68–71				
SM low carbon	4–6	4–6	1.25	62–65				
Ferromolybdenum								
High carbon		1.5	2.5		55–70			
Ferrovanadium								
High carbon		13.0	3.5			1.5		30–40
Ferrotitanium								
Low carbon		3–5	0.1			6–10	38–43	

* In all cases the balance is Fe, with the exception of minor impurities The latter are usually specified, such as 0.10% Max P.

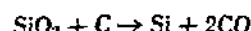
† In several specified grades within this range.

service application, but they are important as the most economical source of these elements for use in the manufacture of the engineering alloys. These same elements can also be obtained, at much greater cost in most cases, as essentially pure metals. The ferroalloys contain significant amounts of iron and usually have a lower melting range than the pure metals and are therefore dissolved by the molten steel more readily than the pure metal. In other cases, the other elements in the ferroalloy serve to protect the critical element against oxidation during solution and thereby give higher recoveries. Ferroalloys are used both as deoxidizers and as a specified addition to give particular properties to the steel (see STEEL MANUFACTURE).

Many ferroalloys contain combinations of two or more desirable alloy additions, and well over 100 commercial grades and combinations are available. Although of less general importance, there are other sources of these elements for steelmaking, such as metallic nickel (Ni), silicon carbide, molybdic oxide, and even mischmetal (a mixture of rare earths). A few typical ferroalloys are shown in the table. The total annual production in the United States was about 1,800,000 tons in 1957 with a value of $440,000,000. The big three, which account for the major tonnage in this class, are the various grades of silicon, manganese, and chromium. For example, 13 lb of manganese are used on the average in the United States for every ton of open-hearth steel produced. The elements supplied as ferroalloys are among the most difficult metals to reduce from their ores.

The most common grade of ferromanganese is a blast-furnace product (standard ferromanganese), the major variation from pig iron production being the use of manganese-rich ore. Other grades of ferromanganese with low C are made by the processes described in the next sections. A low-Si ferroalloy can be made in the same way, but the tonnage grades are made by other methods.

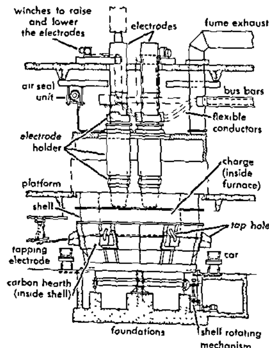
The most general method of ferroalloy manufacture is the submerged-arc furnace (see illustration). Its use will be described in connection with the production of 50% ferrosilicon. A modern furnace of this type is 26 ft in diameter and 10 ft deep with the three carbon electrodes supplying an average of 5 kw-hr per pound of silicon produced. The furnace produces about 2 tons of alloy per hour. The furnace is charged intermittently from the top with a mixture of the required amounts of SiO_2 (quartzite rock), C (coke), and Fe (as turnings), and the ferrosilicon is periodically tapped from the bottom. The over-all operation is essentially continuous, like that of a blast furnace, but the heat is supplied by the resistance of the charge and some arcing, and the coke serves primarily as a reducing agent. After solidification in flat molds, the ferroalloy is crushed to specified sizes for delivery. The over-all reaction for the process can be written



$$SiO_2 + C \rightarrow Si + 2CO$$

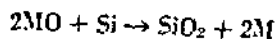
The Si reduced in this way dissolves in the molten iron which is also present and simplifies the operation of the process.

Because SiO_2 is present as gangue in many ores, and because silicate slags have sufficiently low melting points to be controlled conveniently, many other ferroalloys contain appreciable amounts of Si when they are manufactured by this same process. The inherent low solubility of C in Si alloys makes this combination desirable when Si is not objectionable. Low-silicon grades can be made.



Submerged-arc furnace used in ferroalloy manufacture (*Lectromelt Furnace Division, McGraw-Edison Co.*)

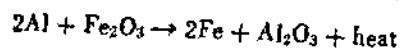
however, by further addition of metal oxide as illustrated by the general reaction



$$2MO + Si \rightarrow SiO_2 + 2M$$

Thus ferrosilicon can be used as an intermediate to produce other ferroalloys, and this scheme is common.

Likewise aluminum can be used as the reducing agent by a process called the thermit reaction



$$2Al + Fe_2O_3 \rightarrow 2Fe + Al_2O_3 + \text{heat}$$

in which all or part of the heat required results from the reduction reaction. The oxide of various desirable alloys may be substituted for Fe_2O_3, and Ca or Mg may be used for Al, a variety of such combinations being possible. The product is low in both Si and C.

Thus, depending upon the reducibility of the ore, and the amounts of C, Si, and Fe which can be tolerated in the product, as well as economic considerations, ferroalloys are produced in the blast furnace, or the submerged-arc or similar variations of the electric furnace, or by aluminothermic or silicothermic reactions. These ferroalloys are then supplied to the steel industry for use as deoxidizers or alloy-addition agents. *See* ARC HEATING; MANGANESE; MOLYBDENUM; SILICON; STEEL; VANADIUM.

[C.D.]

Bibliography: AIME, *Proc. Conf. Electric Furnace Steel*, 1958; Physical Chemistry of Steelmaking Committee, *Basic Open Hearth Steelmaking*, AIME, 1951; B. D. Saklatwalla, Thermal reactions in ferro-alloy metallurgy, the basis of alloy steel development, *Trans. Electrochem. Soc.*, 84: 13–32, 1943; E. E. Thum, Ferrosilicon manufacture at Marietta, *Metal Progr.* 70(4):65–72, 1956.

Ferrocyanide

A compound containing the complex ion, $[Fe(CN)_6]^{4-}$, where iron has an oxidation state of 2+. The coordination complex can be considered a derivative of ferrous and cyanide salts as it is manufactured from these compounds, or a derivative of ferrocyanic acid, $H_4[Fe(CN)_6]$, which has been isolated as a white solid.

Using the nomenclature of coordination compounds, $K_4[Fe(CN)_6]$, for example, should be called potassium hexacyanoferrate(II).

The alkali and alkaline-earth metal hexacyanoferrates(II), with the exception of barium, are easily soluble in water. Most of the heavy metal salts are insoluble; hence, they are used in analytical chemistry.

The sodium and potassium salts are the most important and are used in making blue pigments, dyes, blueprint paper, and ferricyanide.

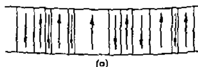
When a solution of ferric salts is added to a ferrocyanide solution, a precipitate with a complicated structure is obtained which is called *Prussian blue*. This precipitate is not $Fe_4[Fe(CN)_6]_3$ which can be prepared under controlled conditions. Prussian blue is now thought to be the same as Turnbull's blue. *See* CYANIDE; FERRICYANIDE; FERROUS COMPOUND; IRON.

[E.E.WB.]

Ferroelectrics

Crystalline substances which have a permanent spontaneous electric polarization (electric dipole moment per cubic centimeter) that can be reversed by an electric field. In a sense, ferroelectrics are the electrical analog of the ferromagnets, hence the name (*see* FERROMAGNETISM). The names Seignette-electrics or Rochelle-electrics, which are also widely used, are derived from the name of the first substance found to have this property, Seignette salt or Rochelle salt. All ferroelectrics exhibit the phenomenon of piezoelectricity, and for this reason they have numerous applications. *See* PIEZOELECTRIC CRYSTAL; PIEZOELECTRICITY.

The reversibility of the spontaneous polarization is due to the fact that the structure of a ferroelectric crystal can be derived from a nonpolarized structure by small displacements of ions. In most ferro-



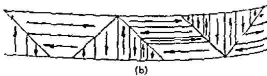


Fig. 1. Domain configurations encountered in ferroelectric crystals (simplified examples). (a) The first class. (b) The second class.

electric crystals, this nonpolarized structure becomes stable if the crystal is heated above a critical temperature, the ferroelectric Curie temperature; that is, the crystal undergoes a phase transition from the polarized phase (ferroelectric phase) into an unpolarized phase (paraelectric phase). The change of the spontaneous polarization at the Curie temperature can be continuous or discontinuous. The Curie temperatures of different types of ferroelectric crystals range from a few degrees absolute to a few hundred degrees absolute. As a rule, the ferroelectric phase is the low-temperature phase; however, there are crystals which are ferroelectric in a relatively narrow temperature range only, and others stay polarized up to the temperature of decomposition or melting.

Classification. Ferroelectrics can be divided into two classes. In ferroelectrics of the first class, spontaneous polarization can occur only along one crystal axis; that is, the ferroelectric axis is already a unique axis when the material is in the paraelectric phase. Typical representatives of this class are Rochelle salt, KH_2PO_4, $(NH_4)_2SO_4$, guanidine aluminum sulfate hexahydrate, glycine sulfate, colemanite, and thiourea.

In ferroelectrics of the second class, spontaneous polarization can occur along several axes that are equivalent in the paraelectric phase. The following substances, which are all cubic above the Curie point, belong to this class: $BaTiO_3$-type (or perovskite type) ferroelectrics; $Cd_2Nb_2O_7$; $PbNb_2O_6$; certain alums, such as methyl ammonium alum; and $(NH_4)_2Cd_2(SO_4)_3$. Some of the $BaTiO_3$-type ferroelectrics have, below the Curie temperature, additional transition temperatures at which the spontaneous polarization switches from one crystal axis to another crystal axis. For example, $BaTiO_3$ and $KNbO_3$ polarize with decreasing temperature first along a [100] axis, then the polarization switches into a [110] axis, and finally into a [111] axis.

Ferroelectric domains. The spontaneous polarization can occur in at least two equivalent crystal directions; thus, a ferroelectric crystal consists in general of regions of homogeneous polarization that differ only in the direction of polarization.

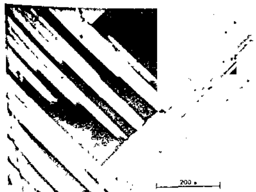


Fig. 2. Ferroelectric domains in $BaTiO_3$ photographed through a polarizing microscope.

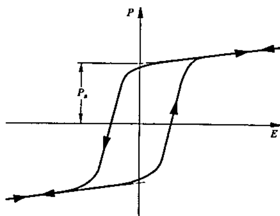


Fig 3. Net polarization P of a ferroelectric crystal vs. externally applied electric field E.

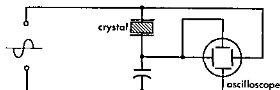


Fig 4. Circuit for the display of ferroelectric hysteresis loops on an oscilloscope.

These regions are called ferroelectric domains. Ferroelectrics of the first class consist of domains with parallel and antiparallel polarization (Fig. 1*a*), whereas ferroelectrics of the second class can assume much more complicated domain configurations (Fig. 1*b*). The region between two adjacent domains is called a domain wall. Within this wall, the spontaneous polarization changes its direction. The wall between antiparallel domains is probably only a few lattice spacings thick, whereas the wall between domains that are polarized at a right angle to each other is probably thicker. Ferroelectric domains can be observed in a number of substances by means of the polarizing microscope (Fig. 2) because of their birefringence, or double refraction (*see* BIREFRINGENCE). The ferroelectric domains range in size from macroscopic (millimeters) to submicroscopic.

Ferroelectric hysteresis. When an electric field is applied to a ferroelectric crystal, the domains that are favorably oriented with respect to this field grow at the expense of the others, for example, by sidewise motion of domain walls. In addition, favorably oriented domains can nucleate and grow until the whole crystal becomes one single domain. When the field is reversed, the polarization reverses through the same processes. The relation between the resulting polarization P of the whole crystal and the externally applied electric field E is given by a hysteresis loop (Fig. 3). The shape of the hysteresis loop depends strongly upon the perfection of the crystal as well as upon the rate of change of the externally applied field E. A [illegible] circuit that permits the observation of ferr[illegible] hysteresis loops by means of an osc[illegible]

shown in Fig. 4. In some ferroelectrics, the polarization can be reversed within a fraction of 1 μsec.

Spontaneous polarization. The magnitude of the permanent or spontaneous polarization P_s of a domain can be obtained from the hysteresis loop by extrapolating the saturation branch to zero external field (see Fig. 3). For most ferroelectrics, the values of P_s are between 10^{-7} and 10^{-4} coulomb/cm² (Fig. 5). In nonferroelectric dielectrics, electric fields between 10^5 and 10^8 volts/cm would be necessary in order to achieve such large polarizations.

Dielectric properties. As a rule, the dielectric constant ε measured along a ferroelectric axis increases in the paraelectric phase when the Curie temperature is approached. In many ferroelectrics, this increase can be approximated by a law of the form

$$\epsilon = \frac{C}{T - T_0} \quad \text{(Curie-Weiss law)}$$

Here T designates the temperature of the crystal, and T_0 is equal to or somewhat smaller than the transition temperature. C is the so-called Curie constant. For $BaTiO_3$, this law holds unaltered up to frequencies of 2.4×10^{10} cps and possible even at much higher frequencies. The dielectric constant drops when the crystal becomes spontaneously polarized (Fig. 6). In the ferroelectric phase, the dielectric constant has two components. The first component is the dielectric constant of the individual domains. It is independent of the frequency and of the electric field generally up to microwave frequencies. The second component is due to domain wall motions, that is, to partial reversal of the spontaneous polarization. This process can give rise to large dielectric losses, and it depends strongly upon the frequency, the electric field strength, the domain structure, and the temperature. In uniaxial ferroelectrics, the dielectric constant measured perpendicular to the ferroelectric axis generally does not show a very pronounced anomaly near the Curie temperature, and in some cases, it has even the same order of magnitude and temperature dependence as for any normally behaving dielectric crystal. *See* CURIE-WEISS LAW; DIELECTRIC CONSTANT; DIELECTRICS.

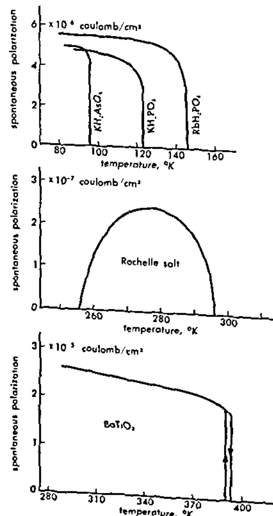


Fig. 5. Dependence upon temperature of the spontaneous polarization of some ferroelectrics.

Piezoelectric properties. Ferroelectrics can be divided into two groups according to their piezoelectric behavior.

The ferroelectrics in the first group are already piezoelectric in the unpolarized phase. Those piezoelectric moduli which relate stresses to polarization along the ferroelectric axis have essentially the same temperature dependence as the dielectric constant along this axis, and hence become very large near the Curie point. The spontaneous polarization gives rise to a large spontaneous piezoelectric strain which is proportional to the spontaneous polarization. In KH_2PO_4-type ferroelectrics and in Rochelle salt, for example, this strain is a shear in the plane perpendicular to the axis of polarization. It reaches 27 min of arc in KH_2PO_4 and about 1.8 min of arc in Rochelle salt. The piezoelectric modulus decreases as the spontaneous polarization increases. However, with sufficiently large stresses, it becomes possible to align the domains and to reverse the spontaneous polarization (Fig. 7). The relation between the resulting polarization of the whole crystal and the mechanical stress is given by a hysteresis loop which is analogous to the loop of Fig. 3 (piezoelectric hysteresis). This effect can simulate a very large piezoelectric modulus.

The ferroelectrics in the second group are not piezoelectric when they are in the paraelectric phase. However, the spontaneous polarization lowers the symmetry so that they become piezoelectric in the polarized phase. This piezoelectric activity is often hidden because the piezoelectric effects of the various domains can cancel. However, strong piezoelectric activity of a macroscopic crystal or even of a polycrystalline sample occurs when the domains have been aligned by an electric field. The spontaneous strain is proportional to the square

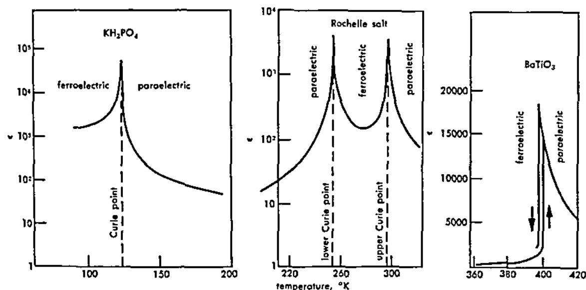


Fig. 6. Anomalous temperature dependence of the relative dielectric constant of ferroelectrics at the transition temperature.

of the spontaneous polarization. In $BaTiO_3$, for example, the crystal (which has cubic symmetry in the unpolarized phase) expands along the axis of polarization and contracts at right angles to it. The strain is of the order of magnitude of 1%. The spontaneous polarization cannot be reversed by a mechanical stress in ferroelectrics of this group.

Crystal structure. The structures of different types of ferroelectrics are entirely different, and it is not possible to establish a general rule for the occurrence of ferroelectricity. The structures of a number of ferroelectrics and the minute changes that they undergo when spontaneous polarization occurs are known in great detail from x-ray diffraction and neutron diffraction studies. In a qualitative way, the process of polarization is best understood for ferroelectrics of the $BaTiO_3$ type. Figure 8 shows schematically the structure of the unit cell of a $BaTiO_3$ crystal in the unpolarized state, and the arrows indicate the direction in which the ions are slightly displaced when the lattice becomes spontaneously polarized along the axis z. The order of magnitude of the displacements is 1% of the unit cell dimension. However, these displacements do not account quantitatively for the observed polarization, because other changes of the electronic structure occur as well. *See* CRYSTAL STRUCTURE.

In KH_2PO_4-type ferroelectrics, hydrogen bonds O—H · · · O play an important part in the ferroelectric effect. Above the Curie temperature, the hydrogen ions are statistically distributed over the two possibilities O—H · · · O and O · · · H—O, whereas below the Curie point, one or the other of these two possibilities is strongly favored, depending upon the sign of the spontaneous polarization.

Antiferroelectric crystals. These materials are characterized by a phase transition from a state of lower symmetry (generally low-temperature phase) to a state of higher symmetry (generally high-temperature phase). The low-symmetry state can be regarded as a slightly distorted high-symmetry state. It has no permanent electric polarization, in contrast to ferroelectric crystals. The crys-

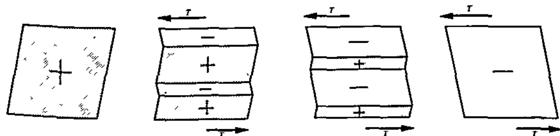


Fig. 7. Reversal of the spontaneous polarization by a mechanical shear stress T in KH_2PO_4 and Rochelle salt (schematic).

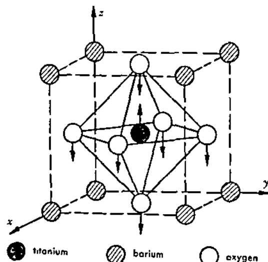


Fig. 8. Crystal lattice of $BaTiO_3$. The arrows indicate the displacements the ions undergo when the crystal becomes polarized.

tal lattice can be regarded as consisting of two interpenetrating sublattices with equal but opposite electric polarization. This state is referred to as the antipolarized state.

In a certain sense, an antiferroelectric crystal is the electrical analog of an antiferromagnetic crystal (*see* ANTIFERROMAGNETISM). In the high symmetry phase, the sublattices are unpolarized and indistinguishable. In general, antiferroelectric crystals have more than one axis along which the sublattices can polarize. Therefore, the low-symmetry

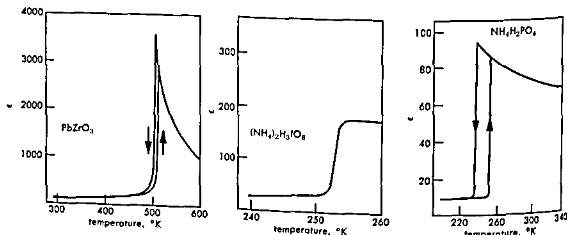


Fig 9. Anomalous temperature dependence of the relative dielectric constant of antiferroelectrics at the transition temperature.

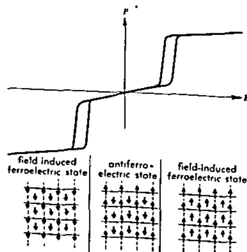


Fig 10 Polarization P of antiferroelectric $PbZrO_3$ vs. externally applied electric field E. Strong fields "switch" the crystal into a ferroelectric state (schematic).

phase consists of regions of homogeneous antipolarization which differ only in the orientation of the axis along which antipolarization has occurred. These regions are called antiferroelectric domains. They can be observed by means of the polarizing microscope. Because these domains have no permanent electric dipole moment, an electric field has, in general, little influence on the domain structure. The dielectric constant of antiferroelectric crystals is generally larger than for nonferroelectric crystals and has an anomalous temperature dependence. It increases as the transition temperature is approached and drops when antipolarization occurs (Fig. 9). In some antiferroelectrics the phase transition is discontinuous; in others it is continuous.

The structure of antiferroelectric crystals is generally closely related to the structure of ferroelectric crystals. Some antiferroelectrics even undergo phase transitions from an antipolarized state into a spontaneously polarized, ferroelectric state; in others, a sufficiently strong electric field applied along an antiferroelectric axis reverses the polarity

of one of the sublattices so that a kind of a ferroelectric state results. The crystal reverts, however, to the antiferroelectric state when the electric field is removed. A plot of the net polarization versus externally applied field for such a case is shown in Fig. 10.

Compounds with antiferroelectric properties are $PbZrO_3$, $PbHfO_3$, $NaNbO_3$ (isomorphous with ferroelectric $BaTiO_3$), WO_3 (structure related to $BaTiO_3$), $NH_4H_2PO_4$ and isomorphous NH_4-salts (isomorphous with ferroelectric KH_2PO_4), $(NH_4)_2H_3IO_6$, $Ag_2H_3IO_6$, and certain alums.

[W.K.]

Bibliography: S. Fluegge (ed.), *Handbuch der Physik*, vol. 17, 1956; D. Megaw, *Ferroelectricity in Crystals*, 1957; F. Seitz and D. Turnbull (eds.), *Solid State Physics*, vol. 4, 1957; A. Van der Ziel, *Solid State Physical Electronics*, 1957.

Ferromagnetism

A property exhibited by certain metals, alloys, and compounds of the transition (iron group), rare earth, and actinide elements in which, below a certain temperature called the Curie temperature, the atomic magnetic moments tend to line up in a common direction. Ferromagnetism is characterized by the strong attraction of one magnetized body for another, a phenomenon known before 600 B.C.

Atomic magnetic moments arise when the electrons of an atom possess a net magnetic moment as a result of their angular momentum. The combined effect of the atomic magnetic moments gives rise to a relatively large magnetization, or magnetic moment per unit volume, for a given applied field. Above the Curie temperature, a ferromagnetic substance behaves as if it were paramagnetic: its susceptibility follows the Curie-Weiss law. The Curie temperature marks a transition between order and disorder of the alignment of the atomic magnetic moments. Some materials exhibit a special form of ferromagnetism below the Curie temperature called ferrimagnetism. *See* FERRIMAGNETISM; *see also* CURIE TEMPERATURE, MAGNETIC; CURIE-WEISS LAW; ELECTRON SPIN; PARAMAGNETISM; SUSCEPTIBILITY, MAGNETIC.

The characteristic property of a ferromagnet is that, below the Curie temperature, it can possess a spontaneous magnetization in the absence of an applied magnetic field. Upon application of a weak magnetic field, the magnetization increases rapidly to a high value called the saturation magnetization, which is in general a function of temperature. For typical ferromagnetic materials, their saturation magnetizations, and Curie temperatures, *see* MAGNETIZATION.

The tasks of a theory of ferromagnetism are to account for the spontaneous magnetization below the Curie point, the temperature dependence of the saturation magnetization, and the nature of the magnetization process, or magnetization curve.

Weiss theory. The Weiss molecular field theory of ferromagnetism (P. Weiss, 1907) represents the first realistic attempt to account for the properties of a ferromagnet. This theory rests on two hypotheses:

1. Below the Curie point, a ferromagnetic substance is composed of small spontaneously magnetized regions called domains. The total magnetic moment of the material is the vector sum of the magnetic moments of the individual domains. It is now known that these assumed domains really exist and are usually between 0.1 and 0.001 cm in thickness.

2. Each domain is spontaneously magnetized because a strong molecular (magnetic) field tends to align the individual atomic magnetic moments within the domain.

The consequence of these assumptions is that while each domain is spontaneously magnetized, the directions of magnetization of the domains do not coincide so that the over-all magnetization of the sample may be much smaller than if it were composed of a single domain. Application of a relatively weak field of the order of 100–1000 oersteds is usually sufficient to align the directions of magnetization of the domains, thereby achieving a large *magnetization.*

The second hypothesis of the Weiss theory leads to the existence of a Curie temperature below which a domain may be spontaneously magnetized in the absence of an applied magnetic field. This comes about in the following way. If the domain is spontaneously magnetized, there must be some sort of interaction between the atomic magnetic moments which tends to align them. Otherwise the domain would behave paramagnetically. This interaction may be represented by an internal magnetic field, the Weiss molecular field, which is proportional to the magnetization of the domain. Thus the field acting on any atomic magnetic moment within the domain may be written

$$H = H_0 + \lambda M \qquad (1)$$

where H_0 is an externally applied magnetic field and λM (M = magnetization) is the Weiss molecular field whose order of magnitude is 10^7 oersteds. It is relatively easy to deduce the magnetic susceptibility above the Curie point since it is known that the ferromagnetic substance behaves like a paramagnet above the Curie temperature. If the Curie law holds,

$$M/H = C/T \qquad (2)$$

where C is the Curie constant. In the Curie law, H is taken to be the local magnetic field acting on an atomic magnetic moment, Eq. (1). From Eqs. (1) and (2)

$$M/(H_0 + \lambda M) = C/T \qquad (3)$$

The susceptibility χ is the magnetization per unit applied field H_0, so that from Eq. (3), in the electromagnetic system of units

$$\chi = M/H_0 = C/(T - C\lambda) = C/(T - T_c)$$

The Curie temperature is defined by $T_c = C\lambda$, and the susceptibility follows the Curie-Weiss law above the Curie temperature. The form of the Curie-Weiss law leads to a nonzero magnetization when $H_0 = 0$ at T_c.

Below the Curie point, where the Curie-Weiss law breaks down, the Weiss theory leads to a spontaneous magnetization and also predicts the temperature dependence of the saturation magnetization. This comes about if it is taken into account that the Curie law is only an approximation valid for weak fields or high temperatures; it does not allow for saturation effects. The correct quantum mechanical expression to replace Eq. (2) is

$$M = NJg\mu_B B_J(a^*) \tag{5}$$

$$a^* = g\mu_B(H_0 + \lambda M)J/kT \tag{6}$$

where N is the number of atoms per unit volume, each with angular momentum quantum number J, g is the spectroscopic splitting factor (the measure of the energy level splittings), μ_B is the Bohr magneton, k is Boltzmann's constant, T is the absolute temperature, and $B_J(a^*)$ is the Brillouin function (*see* GYROMAGNETIC EFFECT; GYROMAGNETIC RATIO; MAGNETON). For $T < T_c$, the value of the magnetization for an applied field of H_0 is obtained by solving Eqs. (5) and (6) simultaneously. In particular, the spontaneous magnetization is obtained by setting $H_0 = 0$ in Eq. (6).

The solution for the spontaneous magnetization ($H_0 = 0$) may be obtained graphically for any temperature T, as shown in Fig. 1. The dashed line of Fig. 1 represents the line $M = a^*kT_c/g\mu_B J$ and is the line for the largest value of T for which there is a solution. This is the Curie point, and for $T > T_c$ there is no spontaneous magnetization. The value of T_c obtained in this way [the slope of Eq. (5) at $a^* = 0$] is $T_c = Ng^2\mu_B{}^2J(J+1)\lambda/3k$, a result which is consistent with the earlier definition $T_c = C\lambda$. The results obtained from Fig. 1 are plotted as a function of temperature in Fig. 2. The value of J chosen is ½.

The plot of Fig. 2 agrees quite well with the experimental results for iron, nickel, and cobalt, thus showing that ferromagnetism is due primarily to electron spin angular momentum ($J = ½$). Since

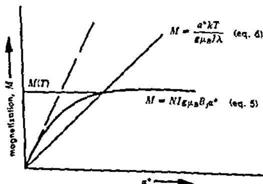


Fig. 1. Spontaneous magnetization below the Curie point.

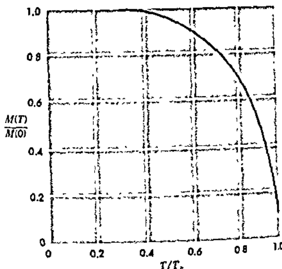


Fig. 2. Temperature dependence of spontaneous magnetization for $J = ½$.

only small fields are required to reach magnetic saturation, Fig. 2 can be expected to give the correct temperature dependence of the saturation magnetization.

Heisenberg theory. The Heisenberg theory of ferromagnetism (W. Heisenberg, 1928) treats the origin of the Weiss molecular field on an atomic basis. It may be remarked at the outset that ordinary dipole-dipole interactions among atomic magnetic moments are much too small to account for the Weiss field. The foundation of the Heisenberg theory is the Pauli exclusion principle. Consider two adjacent atoms of a ferromagnetic material. Each has an atomic magnetic moment arising from electron spin angular momentum. The Pauli principle requires that electrons with parallel spin on adjacent atoms keep out of each other's way. If electrons with parallel spin are farther apart than they would be if their spins were antiparallel, then the electrostatic Coulomb repulsion between them will be less in the former case. The lowest energy state of the system will thus occur when the electron spins are parallel, and this arrangement will be favored. However, if the electrons are kept apart, their kinetic energies will increase and this will tend to make the antiparallel arrangement the state of lowest energy. It is the competition between these two effects which determines whether or not a substance is ferromagnetic. Heisenberg showed that the whole situation could be expressed in terms of an effective interaction between the electron spins, the exchange interaction energy, which had first been derived by P. A. M. Dirac:

$$\text{Exchange energy} = -2J_{ij}\mathbf{S}_i \cdot \mathbf{S}_j \tag{7}$$

Here $\mathbf{S}_i$ and $\mathbf{S}_j$ are the spin angular momentum vectors of the two electrons i and j, and J_{ij} is the so-called exchange integral between the electrons. The exchange integral depends in a complicated way upon the spatial distribution (wave function) of the electrons and is extremely difficult to com-

pute. If the exchange integral is positive, the parallel arrangement is favored, and if J is large enough, ferromagnetism will result. If J is large and negative, antiferromagnetism or ferrimagnetism arises (*see* ANTIFERROMAGNETISM). Arguments can be given which lead to the expectation that J will be large and positive for the interaction between electrons in unfilled d and f shells of atoms of the iron and rare earth groups respectively. The order of magnitude of J is given by $J \sim kT_c \sim 10^{-13}$ erg. *See* EXCLUSION PRINCIPLE.

There seems to be no question that the Heisenberg theory correctly accounts for the tendency of electrons in the same ferromagnetic atom to exhibit parallel spins. However, whether or not it leads to the correct explanation of the interatomic alignment of spins is still a subject of much controversy. There is general agreement that if the Heisenberg theory is the correct approach, then the simple form of the theory is insufficient. In any case, there is no doubt that the interactions which cause ferromagnetism are exchange interactions of the sort discussed by Heisenberg.

Crystalline anisotropy energy. This accounts for the experimental fact that ferromagnets tend to magnetize along certain crystallographic axes, called directions of easy magnetization. For example, a single crystal of iron, which is made up of a cubic array of iron atoms, tends to magnetize in the directions of the cube edges. It requires about 1.4×10^5 ergs/cm³ (at room temperature) to move the magnetization into a hard direction along a cubic body diagonal. Single-crystal magnetization curves, for various directions of applied field H, are shown in Fig. 3.

N. S. Akulov (1929–1931) showed that the anisotropy energy U_A could be expressed conveniently in an ascending power series of the direction cosines between the magnetization and the crystal axes. For cubic crystals, the lowest-order terms take the form

$$U_A = K_1(\alpha_1^2\alpha_2^2 + \alpha_2^2\alpha_3^2 + \alpha_3^2\alpha_1^2) + K_2(\alpha_1^2\alpha_2^2\alpha_3^2) \quad (8)$$

where α_1, α_2, α_3 are direction cosines with respect to the three cube edges, and K_1 and K_2 are temperature-dependent parameters characteristic of the material, called anisotropy constants. In general, $|K_1| > |K_2|$, and further terms are unnecessary. In iron, K_1 is positive, and therefore U_A is a minimum when any single direction cosine $\alpha_i = 1$ and the other $\alpha_j = 0$. That is, the cube edges are easy directions. In nickel, K_1 is negative, and hence the body diagonals are easy directions.

For crystals of other than cubic symmetry the energy U_A must be expressed in a form different from Eq (8) and appropriate to the particular symmetry.

Anisotropy constants can be determined from (1) analysis of magnetization curves, (2) the torque on single crystals in a large applied field, and (3) single-crystal ferromagnetic resonance (*see* MAGNETIC RESONANCE).

The Heisenberg exchange energy, Eq. (7), is isotropic and cannot account for the observed anisotropy, which probably has its origin in a complicated interplay of spin-orbit coupling, crystalline electric fields, and overlap of orbital wave functions. Anisotropy energy depends on the state of strain of the crystal, giving rise to magnetostriction, that is, changes in length of a substance when it is magnetized (*see* MAGNETOSTRICTION).

taneous magnetization, formed at temperatures below the Curie point, are known as domains. As shown in Fig. 4, domains originate in order to lower the magnetic energy. In Fig. 4*b* it is shown that two domains will reduce the extent of the external magnetic field, since the magnetic lines of force are shortened. On further subdivision, as in Fig. 4*c*, this field is still further reduced.

An alternate way to describe the energy reduction is to note that the interior demagnetizing fields, coming from surface poles, are much smaller in the long, thin domains of Fig 4*c* than in the "fat" domain of 4*a*

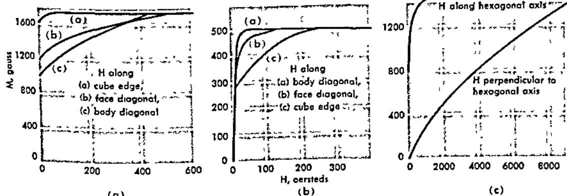


FIG. 3. Magnetization curves for single crystals of (a) iron, (b) nickel, and (c) cobalt. (*After K. Honda and S. Kaya, 1926–1928*)

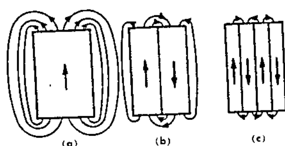


Fig 4. Lowering of magnetic field energy by domains. (a) Lines of force for a single domain. (b) Shortening of lines of force by division into two domains. (c) Reduction of field energy by further subdivision.

The question arises as to how long this subdivision process continues With each subdivision there is a decrease in field energy, but there is also an increase in Heisenberg exchange energy, since more and more magnetic moments are aligning antiparallel. Finally a state is reached in which further subdivision would cause a greater increase in exchange energy than it would cause decrease in field energy, and the ferromagnet will assume this state of minimum total energy.

Bloch wall. Also because of exchange energy, the reversal of magnetization between domains does not occur abruptly, but takes place gradually through a transition zone called the Bloch wall (F. Bloch, 1932). See Fig. 5. To understand the reason for this wall, consider the exchange energy involved. Let the angle between the magnetization of neighbor planes be ϕ. According to Eq. (7), the exchange energy between atoms on neighbor planes varies as $-\cos\phi$, or as a constant $+\phi^2/2$ if ϕ is small. If total reversal takes place in N planes, $\phi = \pi/N$, and the total exchange energy of the wall will be

$$U_E(\text{wall}) \propto \text{constant} + N\tfrac{1}{2}(\pi/N)^2 \qquad (9)$$

The question now arises as to why N should not continue to increase until the entire crystal becomes a single Bloch wall. This does not happen because the decrease in exchange energy is accompanied by an increase in anisotropy energy. Many of the intermediary planes of Fig. 5b must of necessity have their magnetization along hard directions. The larger the value of N, the greater must be the number of such planes. The thickness of the wall can be determined by finding that value of N which makes the sum of exchange and anisotropy energies a minimum. In iron the wall is ~500 angstroms thick and has the total energy ~2 erg/cm².

Domain arrangement. The orientation of domains in a crystal is determined primarily by the need to minimize the magnetic energy (Fig. 4). It is possible to eliminate all surface magnetic poles by forming flux-closure domains (Fig. 6). Here the normal component of magnetization is continuous across all domain boundaries. The demagnetizing fields are zero everywhere, except for the trivial effect of surface poles in the Bloch walls. In a uniaxial crystal, that is, a crystal with a single easy direction, arrangement 6b will be preferred since it has a lower density of magnetization normal to the easy direction, or in the hard direction. Even in cubic crystals, where all directions of magnetization in Fig. 6 may be easy, 6b will be preferred because of magnetostriction. In iron, for example, each domain increases in length along the direction of magnetization by a fraction $\sim 2 \times 10^{-5}$. Thus the domains of Fig. 6 can be fitted smoothly together only by straining them elastically against this magnetostriction, and the required elastic strain energy will be smaller in 6b than in 6a.

In some uniaxial crystals, for example cobalt, the anisotropy is so large that even an arrangement like 6b has too great an energy. No flux-closure domains at all are formed, and the domain pattern is that of Fig. 4c.

In polycrystals the domain structure is much more complicated, depending upon such variables as grain orientation and grain boundaries. It is possible, however, for domains to cross grain boundaries. *See* GRAIN BOUNDARIES (METALLURGY).

On minimizing the total contributions from (1) magnetic, or demagnetizing energy, (2) anisotropy, (3) magnetostriction, (4) elastic strain, and (5) Bloch wall energy, it is found that, depending upon the composition and shape of the crystal, the theoretical domain thickness should vary from about 0.1 to 0.001 cm.

Bitter powder patterns. Direct experimental evidence of the existence of domains is furnished by Bitter powder patterns (F. Bitter, 1931). First a surface is prepared by careful polishing with fine abrasive, followed (for metallic samples) by electropolish. Then a drop of a colloidal suspension of magnetic Fe_2O_3 particles is placed on the surface

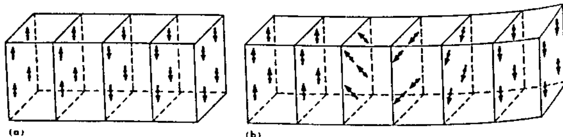


Fig 5. Lowering of exchange energy by the Bloch wall. The reverse of magnetization between domains does not take place abruptly as shown in (a) but by degrees as in (b).

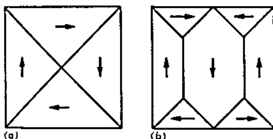


Fig. 6. Flux-closure domains, showing two possible arrangements. (a) Large domains at right angles. (b) Reduction in size of right-angle domains, causing reduction of anisotropy energy of uniaxial crystals or of strain energy of cubic crystals.

and covered by a microscope cover glass. These particles gather at surface regions where the magnetic field is largest, and can be observed easily with a microscope. It is seen from Fig. 5 that the surface field is largest at the center of the Bloch wall; that is, lines of surface poles will demarcate the wall. Thus the colloidal particles concentrate along domain boundaries, as shown in Fig. 7. If the surface deviates by even a few degrees of angle from a simple lattice plane, the pattern becomes very complicated. The beginnings of this may be seen in the "fir tree" patterns at the upper left of Fig. 7. The plane with Miller indices (100) is shown in the figure. For explanation of Miller indices, *see* CRYSTALLOGRAPHY.

The powder patterns only delineate the boundaries and do not themselves show the direction of magnetization. This direction may, however, be inferred from a study of how particles gather around a scratch made on the surface, coupled with observation of domain growth when the sample is placed in a magnetic field.

The Kerr magnetooptic effect (rotation of the plane of polarization of light reflected from a magnetic surface) has also been used to study domains. J. F. Dillon has shown that single-crystal slabs of ferromagnetic yttrium iron garnet (YIG) are transparent to visible light, and that the Faraday effect (rotation of the plane of polarization of light transmitted along a magnetic field) can be used to observe domains. The effect is very striking, and the growth and diminution of domains are readily visible.

Bloch wall motion. This accounts for the initial portion of the magnetization curve. As the wall passes through regions of crystal strain, or over a number of foreign atoms, it may suddenly go over a potential energy hump and into a minimum. There will be a small, almost discontinuous, jump in the magnetization, easily detected by a search coil (a device for measuring change of flux density) and amplifier. This is called the Barkhausen effect. It was once thought erroneously that this effect comes from a sudden complete reversal of magnetization within a domain. However, H. J. Williams and W. Shockley have shown that many Barkhausen jumps accompany the motion of a single domain wall.

Wall motion is hampered by (1) large internal strains and (2) inclusion of nonmagnetic material. The effect of (1) may be minimized by careful annealing and by changing sample composition in order to reduce magnetostriction. The effect of (2) may be minimized by control of sample purity as well as by changing the sample composition to reduce the anisotropy energy. The latter reduction decreases the wall energy and hence smooths out the potential energy humps involved in passing the wall across inclusions. L. Néel has shown that the demagnetizing fields at inclusions cause small flux-closing domains to form around the inclusions, and it is these which are primarily responsible for the potential humps in the wall motion across inclusions. For a discussion of the role of wall motion in the magnetization process, *see* MAGNETIZATION.

Magnetic materials. Much attention has been given to producing better magnetically soft (easily magnetized) materials by reducing anisotropy and magnetostriction. J. L. Snoek (1947) produced a mixed ferrite, called Ferroxcube, with essentially no anisotropy or magnetostriction. Metal alloys with small anisotropy, such as permalloy, have maximum initial permeability at the composition corresponding to zero magnetostriction, as shown by R. M. Bozorth (1953). *See* FERRITE; MAGNETIC MATERIALS, PERMALLOY

Particle shape anisotropy. Small particles have interesting magnetic properties, especially when precipitated within an otherwise nonmagnetic alloy, as in Alnico 5. L. Néel has shown that below a critical radius ($\sim 10^{-6}$ cm for iron) it pays energetically for small spheres to become single domains, and C. Kittel has shown that this is also

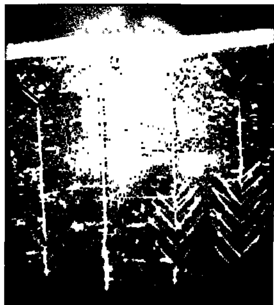


Fig. 7. Bitter powder patterns on a (100) surface of silicon-iron. (*Photograph by H. J. Williams*)

true for thin films. Single domains orient in an applied field by rotation of the magnetization against anisotropy. Néel in particular has discussed the effect of particle shape anisotropy in elongated particles which arises from changes in the demagnetizing fields as the particles rotate with respect to the applied field. Alnico 5 is produced by heat treatment in a magnetic field, which seems to favor the nucleation of platelet-type particles with large shape anisotropy. H. J. Williams and E. A. Nesbitt have shown that this shape anisotropy is the source of the large coercive force in Alnico 5. *See* ALNICO. [E.A.; F.K.]

Bibliography: R. M Bozorth, *Ferromagnetism,* [illegible] *Solid State Physics* [illegible] *magnetic Domains,* 1954; J H Van Vleck. A survey of the theory of ferromagnetism, *Revs Modern Phys.,* 17:27–47, 1945.

Ferrous compound

A compound containing iron in the 2+ oxidation state. This oxidation state, iron(II), is obtained when iron metal is dissolved in nonoxidizing acids such as hydrochloric or dilute sulfuric. Insoluble ferrous hydroxide, $Fe(OH)_2$, is a white, gelatinous solid which turns green, then reddish-brown, as it is oxidized to the ferric state on exposure to air. Solutions of ferrous salts have a pale green color and are also oxidized by air to ferric salts.

Ferrous sulfate is obtained as a by-product from the pickling of steel and by air oxidation of pyrites, FeS_2, and is crystallized from water as greenish crystals, $FeSO_4 \cdot 7H_2O$. It is used in water purification, writing inks, pigments, and medicine.

Mohr's salt or ferrous ammonium sulfate, $Fe(SO_4) \cdot (NH_4)_2SO_4 \cdot 6H_2O$, crystallizes as light-green crystals from solutions containing ferrous sulfate and ammonium sulfate. Because the solid is not easily oxidized by air, it is used in the analytical laboratory as a source of ferrous ion. *See* FERRIC COMPOUND; IRON. [E.E.WR.]

Fertilization

The series of physicochemical processes involved in the union of the male and female sex cells or gametes, the spermatozoon and the ovum, to form a single cell, the zygote. Two principal phases of fertilization may be distinguished: activation of the egg by the sperm to initiate development, and syngamy, or union, of the male and female pronuclei to recombine the respective hereditary contributions in the new individual. Fertilization occurs throughout both the animal and the plant kingdoms and is a fundamental characteristic of bisexual, or bigametic, reproduction. In dioecious organisms the sexes are usually separate, the male organs being in one individual and the female organs in another. In certain invertebrate animals and in many plants, both types of sex organs are present in the same individual, which is then described as a monoecious, or hermaphroditic, organism. In some such forms, like the tapeworm, self-fertilization occurs; in others, like the earthworm, this is prevented by structural features and cross-fertilization takes place.

Fertilization occurs internally in all higher animals including reptiles, birds, and mammals and also in many less evolved forms, such as certain mollusks, insects, fish, and amphibians. The spermatozoa are transferred from the male to the female genital tract during copulation and are carried to the site of fertilization, where gametic union takes place. The sperm are usually transferred as highly active, free-swimming cells in the seminal plasma, which is derived from accessory sex glands such as the seminal vesicles and the prostate. Occasionally, however, the sperm are introduced into the female animal in bundles or packets, called spermatophores, and are freed as required only after a period of quiescence. Insemination, or the natural transfer of spermatozoa during copulation, can sometimes be replaced by artificial means. This practice is of great value in animal husbandry and of some clinical importance in human fertility. When fertilization occurs externally, the more common type of reproduction among aquatic animals, the sperm and the eggs are shed in prodigious numbers, and spawning is synchronized so as to ensure union of the gametes. The meeting of the egg and the sperm is largely at random; only in certain ferns and algae is there an attraction of the sperm by the egg (chemotaxis). Fertilization in the seed plants is facilitated by the enclosure of the sperm nucleus in the pollen tube which grows down the pistil toward the ovule and carries the male nucleus to the ovum.

Variations in fertilization. Variations in the normal process of fertilization are known to occur naturally and may be induced by artificial means. Parthenogenesis, the development of an egg without benefit of activation by, or union with, the sperm, occurs routinely among certain groups of animals, for example, in certain insects. It has also been reported among vertebrates, including fish and birds, and can be induced by artificial activation of the eggs of frogs, salamanders, and rabbits. Gynogenesis is the development of the egg, following activation by the sperm which, however, plays no further role in the process of development. This occurs normally in the nematode worm, *Rhabditis,* for example, and can be simulated in certain echinoderms and amphibians by activating eggs with intensely irradiated, and therefore damaged, spermatozoa. Conversely, the eggs of certain species may be irradiated and then fertilized by normal sperm; the resulting development involves only the male nucleus and is known as androgenesis. Indeed, an egg fragment containing no nucleus can be induced to develop partially after activation by normal sperm; this is merogony. The penetration of the egg by more than one sperm, or polyspermy, occurs normally in some amphibians, birds, and invertebrate animals, but in these cases, only one sperm pronu-

cleus fuses with the egg pronucleus, the other sperm degenerating in the egg cytoplasm. *See* EMBRYOLOGY, EXPERIMENTAL.

Fecundity. Whereas multiple, simultaneous fertilization by a number of sperm of many eggs released at ovulation (superfecundation) occurs among most animals, fecundity is restricted in many birds and mammals so that only one fertilization event occurs at each reproductive period. In women, one ovulation normally occurs, and one fertilizable ovum is produced, approximately each month; exceptional cases may result in multiple ovulation, fertilization, and birth. Identical twinning, however, results from the unusual separation of the single germinal mass, with the result that the sex and the genetic characteristics of the twins are alike

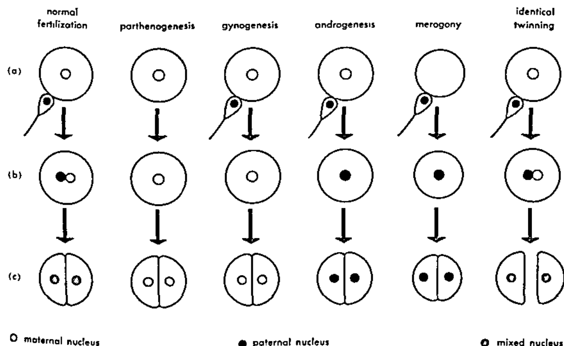
The ovum. The size of the living egg is surprisingly constant, regardless of the size of the animal or the numbers produced. The starfish egg and the human egg are approximately 100 μ in diameter, barely visible to the naked eye. The accumulation of the nutritive yolk, however, may increase the total bulk to considerable size, as in the egg of the domestic fowl. Sperm size, too, is relatively constant but extremely minute in comparison with the egg—one ten-thousandth or less in size and weight. Despite size differences, they both contain the hereditary contributions and physiological mechanisms necessary to initiate a new individual. *See* OVUM.

The union of the sperm and the egg at fertilization reconstitutes the full set of hereditary material, inasmuch as the number of chromosomes has been reduced by one-half during the formation of each gamete. The process of chromosome reduction, meiosis, is completed while the sperm are still within the testis, the male genital organ. In the ovary, the meiotic process of the egg cell need only have begun, since the mammalian egg, at rest after ovulation, completes the process of chromosome reduction upon stimulation by the sperm.

Process of fertilization. The main events of fertilization are best known from studies of marine invertebrate eggs. External fertilization facilitates such a study and permits close observation of the process. Fertilization in vitro has never been satisfactorily investigated in mammalian eggs. The sequence of the main events of fertilization as it occurs in the egg of the sea urchin, an echinoderm, may be presented in the following manner. *See* INVERTEBRATE EMBRYOLOGY.

1. Sperm approach. This is mainly a random chance affair ensured by the presence of vast numbers of gametes.

2. Sperm activation. An increase in the motility and metabolic activity of spermatozoa occurs, induced by substances which exude from the egg surface, specifically from the jelly coat. The mucopolysaccharide fertilizin, derived from the egg, plays an active role in this phenomenon of stimulation and may also act as a recognition signal between sperm and eggs of the same species since it shows serological relationships with the complementary substance, antifertilizin, present in the sperm. Further, fertilizin causes significant structural alterations in



Normal and modified fertilization processes. Stages represented: (a) egg and sperm contact; (b) pronuclei; (c) first cleavage.

the head of the sperm, which initiates physical contact with the egg. The anterior cap of the sperm, the acrosome, breaks down in the presence of fertilizin, and projects a filament capable of penetrating the jelly coat and the surface membranes of the egg. The response of the egg is to produce an eruption, the fertilization cone, through which the sperm passes into the egg.

3. Sperm penetration. The entire motile sperm enters the egg, aided by its own swimming movements, enzymatic action, and possibly physical forces within the egg itself.

4. Egg activation. Contact of the acrosome filament initiates important changes in the surface and subsurface layers of the egg. This is noted as a propagated wave which passes around the egg, beginning at the point of sperm contact. The changes include a breakdown and migration of granular and vesicular components in the egg-cell cortex which may be observed microscopically, and alterations in the electrical, permeable, and viscous properties which have been measured. These rapidly occurring modifications are accompanied, on the one hand, by the initiation of important enzymatic and metabolic processes within the egg and, on the other hand, by the formation of a new surface layer, the fertilization membrane. This membrane, which requires 5–10 sec for its formation, is apparently not a mechanical barrier to the entrance of supernumerary sperm.

5. Pronuclear changes. The sperm nucleus rapidly undergoes changes in preparation for its union with the female pronucleus. Meanwhile, the first division spindle is formed between the centrioles derived from the sperm. The chromosomes of the male and female pronuclei come together on the spindle, forming a mitotic figure which culminates in the first cleavage division. The tail and midpiece of the male gamete degenerate and are absorbed.

The striking and significant aspect of fertilization is the fact that these two relatively small cells selectively combine to produce a single cell ...

between the sperm and the egg, the rapid penetration of the egg by the sperm, the stimulation of physicochemical changes in the egg, and the mutually coordinated interaction of male and female genetic and biochemical activities within the zygote which lead to development. *See* ANDROGENESIS; GYNOGENESIS; MEROGONY; PARTHENOGENESIS.

[D.W.B.]

Bibliography: D. W. Bishop and R. J. Blandau, *Sex and Internal Secretions*, 3d ed., 1959; N. M. V. Rothschild, *Fertilization*, 1956; A. Tyler, *Analysis of Development*, 1955; E. B. Wilson, *The Cell in Development and Heredity*, 3d ed., 1925.

Fertilizer

Natural or manufactured products used to promote plant growth by supplying essential chemical elements. The elements supplied in greatest quantity by fertilizers are nitrogen, phosphorus, and potassium. These elements are not available to higher plants in elemental form and are supplied in chemical combination with one or more other essential ...

ganese, molybdenum, and zinc) are supplied in smaller quantities by fertilizers.

Approximately 25% of the annual production of food and fiber in the United States is attributable to increases in crop yields resulting from the use of commercial fertilizer.

NATURAL FERTILIZERS

Organic materials derived from plant and animal sources, including farm manures, farm wastes, various seed meals, fish and animal tankages, and sewage sludges, are widely used. Their contents of nitrogen and other essential nutrients are low, and they are valued largely because of their low solubility, resistance to leaching, and tendency to improve the texture and increase the biological activity of the soil. Typical analyses of some natural organics marketed for fertilizer use are given in Table 1. *See* MANURE.

Table 1. Typical analyses of some natural organic fertilizers

Material	Nitrogen content, %	P_2O_5 content, %	K_2O content, %
Animal tankage	5.7–10 0	1.8–3 6	0 1–1 6
Dried blood	12.0–14 0	0 5–2 0	0 1–0 9
Castor pomace	4 0–6 5	1 0–2 0	1 0–1 5
Cocoa tankage	2 0–2.5	1 0–1 3	0 6–3 0
Cottonseed meal	6.5–7 5	2.0–3 0	1.5–2 0
Fish tankage	6 5–10.0	4 0–8 0	0 1–1 1
Garbage tankage	2.5–3 3	2 0–5 0	0 5–1 0
Tung meal	3.8–4 4	1.0–1.5	1 0–1 5
Sewage sludge	1.5–6 3	1.0–4 0	0 1–0 5

Compost is usually low in nutrient element content. It is prepared by decomposition of farm and garden wastes such as straw, corn stalks, corn cobs, vines, leaves, garbage, or practically any plant or animal tissue. Composting is employed to avoid decomposition in the soil by microorganisms which consume nitrogen and other nutrient elements and decrease the supplies readily available for plant growth.

Methods of composting vary widely, but in all cases a pile of the organic material is kept moist to maintain conditions favorable for decomposition and evolution of carbon dioxide. Frequently, ammonium salts are added to accelerate decomposition, phosphate to improve quality, and limestone to neutralize acids formed. The compost is suitable for general use when the material has decomposed so that it possesses no objectionable odor and is brown to black in color. It will usually have lost 50% or more of its original weight and will contain 1.5–2.5% of nitrogen.

MANUFACTURED FERTILIZERS

Essential nutrient element compounds and their mixtures are derived from atmospheric nitrogen,

mineral deposits, saline brines, or industrial operations and are marketed as fertilizers.

Nitrogen fertilizers. These include sodium nitrate from Chilean deposits, ammonia liquor and ammonium sulfate (by-products from the coking of coal), and products prepared from atmospheric nitrogen by the cyanamide and synthetic ammonia processes.

In the cyanamide process, in use in Europe, Asia, and Canada, coke and limestone react in an electric furnace to produce calcium carbide, which absorbs gaseous nitrogen at 1000°C to form calcium cyanamide, $CaCN_2$. Formerly, appreciable quantities of cyanamide were used to produce ammonia, ammonium sulfate, and urea.

The synthetic ammonia process, widely used throughout the world, produces ammonia from 3:1 mixtures of hydrogen and nitrogen at temperatures of 425–650°C and pressures of 100–1000 atm in the presence of a promoted iron catalyst. In some cases, hydrogen for the synthesis gas is produced by electrolysis of water, or in the electrolytic production of caustic soda; nitrogen is produced by fractional distillation of liquid air. More frequently, the synthesis gas is obtained by reaction of steam and air with coal or coke, or with natural gas, petroleum refinery gases, or other hydrocarbons in accordance with producer and water gas reactions. The carbon monoxide formed is catalytically converted to carbon dioxide, and with other impurities, is removed from the system. *See* AMMONIA.

The synthetic ammonia process is far more important than the cyanamide process because the cost of nitrogen fixation is lower, and the ammonia may be used directly as fertilizer or be readily processed into ammonium nitrate, ammonium sulfate, sodium nitrate, urea, and other nitrogen compounds. In 1959, the United States industry included 56 plants with a total annual productive capacity for both agricultural and industrial use of 4,120,000 short tons of nitrogen. This was based almost exclusively on natural gas and petroleum refinery gases as the hydrogen sources. Typical compositions of chemical nitrogen fertilizers are given in Table 2.

Table 2. Typical analyses of chemical nitrogen fertilizers

Material	Nitrogen content, %
Ammonium nitrate	33 5
Ammonium nitrate-limestone mixtures	20 5
Ammonium phosphate*	11 0–21 0
Calcium cyanamide	21 0
Calcium nitrate	15 5
Sodium nitrate	16 0
Urea	45 0
Anhydrous ammonia	82 0
Aqua ammonia†	20 0–25 0
Nitrogen solutions‡	33 5–49 0

* Available P_2O_5 content, 20 0–53 0%
† Aqueous solution of ammonia
‡ Aqueous solutions of ammonia with ammonium nitrate or urea, or both

Ammonium nitrate, ammonium phosphates, and ammonium sulfate are prepared by absorption of ammonia in liquid nitric, phosphoric, and sulfuric acids or mixtures of these acids. Synthetic sodium nitrate, in contrast to the natural Chilean product, is produced by neutralization of nitric acid with sodium carbonate, or the reaction of nitric acid with salt to yield chlorine as a coproduct. Nitric acid for these purposes is prepared by catalytic oxidation of ammonia with air and subsequent absorption of the nitrogen oxides in water.

Urea, $CO(NH_2)_2$, is produced by reaction of ammonia and carbon dioxide at 160–200°C and 100–400 atm pressure. Numerous modifications of the basic process are in use, but all rely on the formation of ammonium carbamate and its decomposition to urea and water. By early 1959, the estimated annual productive capacity of the 12 direct synthetic urea plants in the United States was 793,000 short tons. *See* UREA.

Aqua ammonia, anhydrous ammonia (a liquefied gas), and the nitrogen solutions (some of which exhibit appreciable vapor pressures), are applied directly to the soil or in irrigation water, and are used in the manufacture of mixed fertilizers. The other materials listed in Table 2 are water-soluble solids and normally are handled in the dry condition. Nitrogen solutions rank first among the sources of nitrogen for mixed fertilizers in the United States, and anhydrous ammonia is one of the principal nitrogen materials used directly as fertilizer

Phosphorus fertilizers. There are three principal sources of phosphorus fertilizers: (1) mineral phosphates or phosphate rock; (2) basic slag, a by-product of the steel industry; and (3) bones, a by-product of the meat-packing industry. Of the three, mineral phosphates containing 25–39% P_2O_5 are by far the most important. *See* APATITE.

Most of the phosphate rock used as fertilizer is processed chemically to improve the utility of the phosphorus for crop production. However, under certain soil conditions, some plants, especially legumes, are able to make good use of the phosphorus in raw phosphate rock.

Superphosphates. Superphosphates, produced by acidulation of phosphate rock, are classified as (1) normal or ordinary, and (2) concentrated—triple, treble, or double—in accordance with their P_2O_5 concentrations and the acids used in their manufacture.

Normal superphosphate, made by treating phosphate rock with sulfuric acid, contains 18–22% available P_2O_5 and is the principal phosphatic fertilizer used in the United States and most other countries. It is essentially a mixture of monocalcium phosphate with small percentages of iron and aluminum phosphates and 50% or more calcium sulfate. Monocalcium phosphate is soluble in water, and iron and aluminum phosphates are mostly soluble in neutral ammonium citrate. Unreacted phosphate and so-called reverted phosphate are insoluble in water and citrate solution, and account

for about 1% P_2O_5. In the United States, chemically processed phosphates and phosphatic mixed fertilizers are sold on the basis of their content of available P_2O_5, which is the sum of the water- and citrate-soluble P_2O_5.

Concentrated superphosphate, made by treating phosphate rock with phosphoric acid, contains 42–50% available P_2O_5, chiefly as monocalcium phosphate. It contains only small amounts of calcium sulfate.

Other phosphates. These are produced from phosphate rock by acidulation and by calcination and fusion processes. Phosphoric acid is produced by both wet and furnace processes. The wet process involves acidulation of phosphate rock with sulfuric acid and separation of the solid calcium sulfate formed from the crude aqueous phosphoric acid. In the furnace process, the phosphate is reduced to phosphorus by carbon in an electric furnace, volatilized, and subsequently oxidized and hydrolyzed to phosphoric acid of high purity.

Processes for the decomposition of phosphate rock with nitric acid usually require supplemental use of sulfuric, phosphoric, or carbonic acids, or their soluble salts, and ammonia. The products, nitric phosphates, contain ammonium nitrate equivalent to the nitric acid used, and the phosphorus occurs largely as dicalcium phosphate with the balance of the calcium present as calcium sulfate or calcium carbonate. Incorporation of potassium salts produces high-analysis complete fertilizers in a continuous operation.

Several high temperature processes for the treatment of phosphate rock to produce citrate-soluble phosphorus have been developed commercially, including heating phosphate rock with (1) sodium carbonate or magnesium silicate minerals, (2) silica and water vapor, and (3) phosphoric oxide (P_2O_5) vapor. The latter process produces a calcium metaphosphate glass containing 63% citrate-soluble P_2O_5.

Basic slag. The second most important source of fertilizer phosphorus is a by-product of the production of steel from high-phosphorus iron ore. It is high in lime and silica and contains moderate amounts of iron, manganese, and magnesium. In slags low in fluorine, the phosphorus content is nearly as readily available as that in superphosphate. European basic slag usually contains about 15% P_2O_5, as compared with 8–10% for the relatively small production in the United States.

Bones. Formerly the chief source of phosphorus, bones now are of lesser importance than either phosphate rock or basic slag. The phosphorus in bones is more completely and readily available to plants than that in phosphate rock. Raw bone meal contains about 22.5% P_2O_5 and 4% nitrogen, and steamed bone meal about 25% P_2O_5 and 2.5% nitrogen.

Potassium fertilizers. These are chiefly water-soluble salts derived from salt-lake and other brines, and in the main, from deposits of water-soluble minerals. The soluble minerals which furnish most of the potassium are carnallite ($KCl \cdot MgCl_2 \cdot 6H_2O$), kainite ($KCl \cdot MgSO_4 \cdot 3H_2O$), sylvite ($KCl$), and langbeinite ($K_2SO_4 \cdot 2MgSO_4$). The ore usually is recovered in the solid state from the deposits by underground mining and is conveyed to the surface for processing. The techniques of concentrating and recovering potassium salts from brines and soluble minerals include flotation, solar evaporation, fractional crystallization and the application of the principles of base exchange, double decomposition, and the phase rule.

The principal potassium fertilizer produced in the United States is high-grade potassium chloride (KCl) or muriate of potash, containing 60% or more potash expressed as K_2O. Next in importance are potassium sulfate (K_2SO_4), containing 48–52% K_2O, and sulfate of potash-magnesia ($K_2SO_4 \cdot 2MgSO_4$), containing 22% K_2O. Also produced are small tonnages of potassium chloride, containing 50% K_2O, and of manure salts (25% K_2O), consisting chiefly of a mixture of potassium chloride and sodium chloride. The sulfates are more expensive than the chloride salts and are used only when the latter are considered unsatisfactory.

Mixed fertilizers. These products contain two or all three of the principal plant nutrients. The composition or grade of mixed fertilizers usually is expressed by three numerals: the first indicates percentage of nitrogen; the second, the percentage of available P_2O_5; and the third, the percentage of K_2O. Most mixtures contain all three nutrients, but some have only two. Those containing all three are commonly called complete fertilizers.

The kinds of mixed fertilizers produced include solids, liquids, powders, and granules, as well as numerous specialty products for nonfarm use. The materials used in their manufacture range from gases and liquids to solids that differ greatly in their physical characteristics. The manufacture of mixed fertilizers in the United States involves chiefly the relatively simple process of screening and blending the required quantities of solid fertilizer materials—usually with the addition of ammonia either as anhydrous ammonia or as an aqueous solution also containing ammonium nitrate or urea—with other materials, such as conditioning and neutralizing agents, and with sufficient filler to adjust the concentration of the total nutrients to the desired level. In this operation, the uncombined or neutralizing ammonia reacts with the superphosphate in the mixture, while the dissolved ammonium nitrate or urea is precipitated in the fertilizer mass and the water is combined largely as water of hydration. The mixtures then are cured in large piles to allow the completion of the chemical reactions between the various constituents. The operations and equipment for making powdery mixtures in this way range from simple types requiring much hand labor to highly mechanized plants.

The production of granulated mixtures requires the use of special processes and equipment. Granulation minimizes segregation of the constituents,

reduces caking, decreases loss by dusting, and facilitates the application of the fertilizer in the field. Many fertilizer materials are marketed wholly or partly in granular form.

Liquid mixed fertilizers, containing two or all three of the principal plant nutrients, appear to have the advantage of lower capital investment and labor costs for their production, greater convenience in handling, and greater uniformity of composition and physical condition than solid products. The production of liquid mixtures equivalent in plant-nutrient content to solid mixtures, however, considerably restricts the choice and availability of nutrient materials to those exhibiting a high degree of mutual solubility in water.

The total content of the principal nutrients (nitrogen, P_2O_5, K_2O) in mixed fertilizers in the United States rose from an average of 19.92% in 1940 to 23.58% in 1950 and 30.26% in 1957–1958

Trace elements. Boron, chlorine, copper, iron, manganese, molybdenum, and zinc are required in more limited quantities than the primary elements (nitrogen, phosphorus, and potassium) or the secondary elements (calcium, magnesium, and sulfur). Deficiencies of both calcium and magnesium are corrected chiefly by applications of limestone and other agricultural liming materials. Large quantities of calcium and sulfur are supplied by superphosphates. Other sources of magnesium and sulfur include magnesium sulfate and sulfate of potash-*magnesia. Magnesium* is also supplied as magnesium oxide.

Frequently, adequate amounts of available forms of the trace elements are present in the soil or in the materials used to correct deficiencies of the primary *and secondary elements. Usually they are not* supplied separately, or except in the case of chlorine, incorporated with other fertilizer materials unless a specific need has been demonstrated. Chlorine is applied to soils principally as a component of potassium fertilizers.

Boron commonly is applied in the form of borax (sodium tetraborate), a water-soluble compound, but limited use is made of slightly soluble or insoluble compounds such as calcium borate and borosilicate glasses. Usual applications of borax range from 5 to 40 lb per acre. The use of boron fertilizers must be controlled carefully, however, for even a small excess is toxic to many plants.

Water-soluble sulfates are the usual sources of copper, iron, manganese, and zinc, but increasing use is being made of chelating agents, such as ethylenediaminetetraacetic acid (EDTA), to maintain these elements in nonionized water-soluble condition and so permit their ready absorption by plants under a wide range of soil conditions. Molybdenum deficiencies are corrected by use of 2–16 oz of sodium molybdate per acre. *See* AGRICULTURAL CHEMISTRY; FERTILIZING; PLANT, MINERALS ESSENTIAL TO; PLANT, MINERAL NUTRITION OF; SOIL. [K.G.C.]

Bibliography: F. E. Bear, *Soils and Fertilizers*, 4th ed., 1953; G. H. Collings, *Commercial Fertilizers*, 5th ed., 1955; K. D. Jacob (ed.), *F... Technology and Resources in the United States*, 1953; V. Sauchelli, *Manual on Fertilizer Manufacture*, 2d ed., 1954; J. W. Turrentine, *Potash in North America*, 1943.

Fertilizing

Provision of additional nutrients to enhance plant growth, particularly to increase that part of the crop intended for harvest. Chemical carriers of plant nutrients are usually known as commercial fertilizers. Farm manures also contain plant nutrients and, in addition, are high in organic matter which serves as a soil conditioner. The application of chemical carriers of plant nutrients to the soil in a systematic manner began about 1850 and marked the beginning of scientific crop production

Value. By the 1950s farmers of the United States were investing $1,000,000,000 annually in fertilizers. The importance of this investment is evident when it is realized that the annual gross value of farm production in the United States amounts to approximately $17,000,000,000. About $3,000,000,000 of this is attributable to fertilizers alone.

Grades. Commercial fertilizers are mainly designed to supply one or more of the three major elements nitrogen (N), phosphorus (P), and potassium (K), in suitable chemical form. In this order (N, P, K) the grades of fertilizers are identified and the numbers indicate the percentage of the total weight of each of the three components. *Hence, the numbers 5–10–10 represent a mixture* containing 5% nitrogen (N), 10% phosphorus pentoxide (P_2O_5), and 10% potassium oxide (K_2O).

Physical forms. Chemical fertilizers are mar*keted in both dry and liquid forms. Dry forms in*clude powdered, granulated, and pelleted fertilizers. Liquid fertilizers are obtainable in high-pressure and low- or nonpressure forms. Mixed fertilizer solutions are usually nonpressure liquids.

Application methods. Methods of applying fertilizers for maximum crop yields vary considerably with such factors as kind of crop, time of application, rate of application, chemical properties of fertilizers applied, and soil conditions. If ample funds are available, large quantities of fertilizer [illegible]

precise equipment for placement, and more care in selection of fertilizer materials. For example, in some exceptional cases, such as high-return vegetable crops, use of continuous bands of fertilizer placed fairly close to plants at the time of planting has resulted in crop production equal to that of *broadcast methods which require three* times the amount of fertilizer. Good response has been obtained on many crops by applying large quantities of starter fertilizer close to seed or plants in moist soil. In this practice, it is important to have ample soil between the seed or roots and the fertilizer to safeguard against injury to th

seedlings or plants. An offset continuous band 1–3 in. to the side and down to 6 in. in depth below seed level is the usual range for efficient use of fertilizer. With transplants, placement of fertilizer usually ranges 3–6 in. to the side of the plant and 2–6 in. in depth.

Equipment. Fertilization equipment has been improved greatly since World War II. Changes have been made to take advantage of new types of fertilizer material, of the higher analyses and higher rates used, and of better methods of application. Rates of application may range from 100 to 3000 lb/acre. There are only a few basic types of mechanical device for metering fertilizers. Dry-fertilizer broadcasters employ either a regulated gate opening over rotating vanes, or an agitating mechanism over an orifice of regulated size. Examples of the first type are truck spreaders and end-gate broadcasters which have effective application widths ranging from a few to 20 ft or more. Some truck spreaders use augers in long distributive tubes and often carry 8–10 tons of fertilizer. Other broadcasters are pull-type machines with widths up to 10–12 ft. Auger agitators are occasionally used on broadcasters. Metering devices have also been adapted to row-crop hoppers and grain drills. The star-wheel dispensing unit is commonly used on drills. This dispenser operates horizontally with fingers carrying increments of fertilizer under a shield which is adjustable in height. It is used on some row-crop fertilizer hoppers, although such machines are generally fitted with a revolving-bottom distributor. The latter type meters fertilizer through an adjustable or a fixed side gate. Some distributors have reciprocating plates on the bottom of the hopper which oscillate slowly or are driven by tractor power take-off at high speeds.

Liquid fertilizer of the high-pressure type, for example anhydrous ammonia, is regulated mainly by a valve or a positive-displacement pump. The size of the orifice may be controlled manually or by a pressure-regulating valve. Special depositors place the liquid into the soil to prevent loss of ammonia. Placement at a 6-in. depth is usually adequate. Low- or nonpressure solutions are usually metered by gravity dribble tubes, compressed-gas systems, gear pumps, roller pumps, metering piston pumps, or metering hose pumps. In the gravity system, orifices of four to six sizes are provided and, by means of an [illegible] (to minimize [illegible] steady flow of [illegible] ompressed-gas [illegible] uses a small compressor to maintain constant pressure in the tank, and is especially efficient for low-pressure solutions because, with a compressor, discharge rates are not affected by temperature changes or changing of void space in the tank. Gear and roller pumps meter nonpressure solutions well when the speed of equipment is kept uniform. The metering piston pumps and metering hose pumps are driven by a ground wheel or comparable drive and provide a constant rate of application regardless of throttle or gear changes. This is desirable with unskilled operators or on fields having irregular surfaces. The metering mechanism is the key part of most field-fertilizing machines. With the exception of broadcasters, the mechanisms are generally used as attachments to other field equipment. Such attachments are used in preplanting operations with subsoilers, plows, harrows, and chisels. Drills for grain, vegetables, or heavy-duty grassland planters or transplanters often have fertilizer attachments, and side- and top-dressing attachments are usually mounted in conjunction with cultivating and spraying equipment. Soil tools (if not a part of the base machine) usually consist of shovels or rolling disks or coulters to effect adequate placement (see illustration).
[W.C.HU.]



Safe placement of fertilizer in reference to seed or seed piece is very important. This picture shows an efficient placement of fertilizer for potatoes. Attachments to make fertilizer placement at the time of planting are available on most field row planters. (USDA Agricultural Research Service)

Effect on growth and yield. For plants to develop, all the nitrogen and mineral elements essential for growth must be supplied by way of the root system or, to a limited extent, through the leaves. The total supplies of mineral elements in certain soils are adequate for many years of crop production. But the rate at which those elements become available for plant use may be too slow. Each nutrient must be in adequate supply and in favorable balance with all the others. Hence potassium, phosphorus, and the several micronutrients may be deficient in rate of supply because of fixation or adsorption by the soil; lack of nitrogen may become a limiting factor through adverse microbial conditions or excessive loss by leaching. Moreover, the rate of supplying nutrients must be favorable throughout the entire growth period.

Fruit or storage organs rather than merely vegetative yield, are often desired. Further, it is frequently the total content of some substance within

the harvested organ that is important rather than the weight or size of the organ itself. Thus, it is the quality of tomatoes or grain or the percentage of sugar in cane or beets that is economically important. The mineral supply at later stages of growth may determine the economic value of the crop. For example, high nitrogen acquisition at later stages of development may lead to increased protein but reduced carbohydrate content. Therefore nitrogen and mineral nutrition must be in favorable accord with development of desirable characteristics of the harvested product.

Fertilizers. For improving the mineral nutrition of plants, favorable combinations of chemicals are compounded and used. These may include all of the mineral elements that are essential to growth. Micronutrients, occurring as impurities in fertilizer-grade chemicals, may be inadequate for maximum plant development, and extra supplies may have to be added. Some latitude is permissible in the balance of elements supplied, but both marked deficiencies and marked excesses of any one element must be avoided.

Commercial fertilizers, manures, and lime or other materials may be added to provide essential mineral elements at required available levels, or to make environmental conditions more favorable, improve structure, or stimulate microbial activity within the soil so that such factors do not become limiting to growth and development. Sometimes reference is made to the use of balanced fertilizers. The balance that is important is not in the fertilizer, but in the soil after the fertilizer has been added and has reacted with the soil. The same fertilizer will produce different effects in different soils. Addition to the soil does not necessarily ensure availability; certain elements, such as potassium, phosphorus, and the micronutrients, are strongly fixed in some soils. This is especially true, where colloidal clay content of the soil is relatively high. *See* AGRICULTURAL MACHINERY; AGRICULTURAL SOIL AND CROP PRACTICES; FERTILIZER; PLANT GROWTH; PLANT, MINERAL NUTRITION OF; PLANT, MINERALS ESSENTIAL TO. [T.C.B.]

Fescue

A group of grasses, of which the most important species is tall fescue (*Festuca arundinacea*). Alta, produced in Oregon, and Kentucky 31 are two improved strains that are similar in appearance and value. Plants grow 3–4 ft tall and are leafy and vigorous. Fescue has a wide range of soil and climatic adaptation and responds to fertile soils in which it develops uniform sods. In mild climates, growth continues throughout winter. Sustained summer growth is typical. Fescue is used for both hay and pasture although it is sometimes unpopular because of the harshness of its older leaves. Good seed habits, ease of establishment, and high productivity have increased its use over much of the United States. *See* GRASS CROPS.

[H.B.S.]

Fetal membrane

An auxiliary organ, more or less membranous in character, lying outside the embryo or fetus proper, hence extraembryonic membrane. Fetal membranes are present in all reptiles, birds, and mammals (amniotes), and certain types of comparable membranes are present in other animals such as the bony fishes and sharks.

Usually four distinct fetal membranes arise during the development and are closely associated with the embryos of reptiles, birds, and mammals. These are the yolk sac, amnion, chorion, and allantois. Certain of the originally separate membranes, like the chorion and allantois in birds, may combine to form a compound structure, the chorioallantois. In higher mammals, the chorion and allantois, or the allantoic blood vessels, may combine with the uterine mucosa, or lining, of the mother to form the placenta. The umbilical cord is formed by the fusion of the amnion, allantois, and yolk sac. Fetal, or extraembryonic, membranes are living, cellulated structures, continuous with specific embryonic body parts or regions from which they have arisen by various processes of folding and evagination. Just prior to or at the time of hatching in reptiles and birds or birth in mammals, the membranes are either absorbed by the embryo, such as the yolk sac in birds, or become separated from it, as do the amnion, chorion, and allantois in some animals, by some natural process of dehiscence. The separated membranes of mammals, mostly the amnion, chorion, and placenta, are collectively termed the afterbirth.

Fetal membranes, in general, serve an indispensable role in the protection, respiration, excretion, and nutrition of the embryo and fetus until such time as the latter is capable of independent exist-

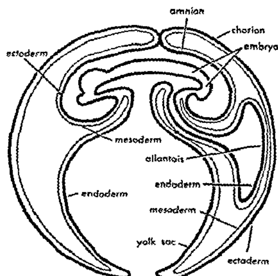


Diagram illustrating the relations of the fetal membranes to one another and to the embryo body.

ence. However, a particular membrane, the yolk sac, may serve an important functional role in some groups, like reptiles or birds, but may exist as little more than a vestigial organ in others, as it is in most mammals. *See* ALLANTOIS; AMNION; CHORION; YOLK SAC. [N.T.S.]

Bibliography: L. B. Arey, *Developmental Anatomy*, 6th ed., 1954; O. E. Nelsen, *Comparative Embryology of the Vertebrates*, 1953.

Fever

Elevation of the body temperature, as a result of disease. The generally accepted upper limit of normal temperature is 98.6°F in the mouth and 99.6°F in the rectum. The body temperature rarely exceeds 106°F per rectum, although brief levels of 112–113°F have been reported in persons who later regained health. In persons with fever, as well as in normal individuals, there is nearly always some daily fluctuation of temperature with higher levels in the afternoon and lower levels in the morning.

Rise of body temperature is associated with many kinds of disease. It is nearly always present in infections, but may also be a prominent manifestation of such disorders as blood diseases, malignant tumors, vascular accidents, or allergies.

Persons with fever are often, but not always, aware of a sensation of warmth. During periods when the internal temperature is rising very rapidly the skin may feel cold and there may be a shaking chill. Other signs and symptoms commonly associated with fever include flushing, sweating, headache, muscle or joint pain, nausea, or vomiting.

The primary mechanism of fever appears to be disturbed function of those centers in the brain (hypothalamus and midbrain) which control mechanisms for heat production and loss. These affect particularly the activity of voluntary muscle fibers and the flow of blood to the surface of the body. The means by which thermoregulation is upset has not been established with certainty, but the most widely accepted hypothesis is that a product of tissue injury affects the function of the cerebral centers. Such activity has been demonstrated in fluids which collect in areas of inflammation, as well as in certain kinds of white blood cells. *See* HOMEOSTASIS. [P.B.BE.]

Bibliography: G. Pickering, Regulation of body temperature in health and disease, *Lancet*, 1(7011):59–64, 1958.

Fiber, man-made

One of a number of textile fibers and filaments manufactured from chemical substances of natural, inorganic, or synthetic origin; the last are also known as synthetic fibers. Although the nomenclature has varied in the past, the correct technical usage makes a distinction based on the origin of the raw material. Accordingly, rayon is a man-made fiber but not a synthetic fiber. Nylon is a synthetic fiber that also falls into the broader classification of man-made fibers. In addition to names based on their chemical composition, man-made fibers are customarily known by the manufacturer's name or trademark, or by terms referring to characteristic properties, for example, thermoplastic fibers, or to the manufacturing process, for example, wet-spun fibers and viscose rayon.

Man-made fibers, with the exception of the inorganic ones, are composed of long, chainlike molecules called linear polymers. The arrangement (orientation) of these polymer chains with respect to the fiber axis, and the manner in which they fit or pack together (crystallinity) is controlled both by their chemical nature and by the manufacturing steps. In turn, these three variables —chemical nature, orientation, and crystallinity— establish the fiber properties such as strength, flexibility, resilience, and durability. Fiber properties are important factors in the appearance, comfort, and wear of garments, home furnishings, and other textile applications.

To spin filaments, the starting material is made fluid by melting or dissolving in a solvent, extruded through small holes, and then solidified by cooling (melt spinning), by chemical treatment (wet spinning), or by evaporation of the solvent (dry spinning). The filaments are produced in various sizes, and they range in diameter from finer than a human hair to large bristles used for toothbrushes.

Filaments are combined to form yarns of two different types: continuous filament yarns and spun yarns. Spun yarns are made from short fibers called staple. These are cut in lengths from 1 to 4½ in., and are held together in the yarn by crimp and twist in a manner similar to that used for natural fibers. The choice of yarn depends on the intended purpose; fabrics of spun-staple yarns are usually more bulky and less smooth and lustrous than fabrics of continuous filament yarns.

Man-made fibers are used both alone and in blends with natural fibers in all types of fabrics for clothing, household, industrial, and military purposes. In the principal textile uses, man-made fibers account for about 23% of the total world consumption, cotton, 66%, and wool, 11%. Man-made fibers are made in 40 countries. The United States leads with a total annual production of about 2,000,000,000 lb. The industry has grown steadily by applying modern science and technology to improve fiber properties and to lower costs.

An understanding of why only certain substances can be drawn out into ductile fibers developed in the period from 1920 to 1930. The classical researches of Emil Fischer and Hermann Staudinger on natural fiber-forming substances such as cellulose, starch, and silk fibroin showed that molecules of tremendous size (macromolecules) are necessary. Early theories were brilliantly confirmed and extended by the work of Wallace Carothers. His research led to the discovery of nylon and also established the principles by which countless numbers and kinds of linear high polymers can be synthesized. Continued research on raw materials in many laboratories has led to an increasing number of possible fiber types, some excellent and some

disappointing. Only a fraction of those tried experimentally or on pilot-plant scale prove to have all of the characteristics necessary for commercial success. The fiber types of major importance in this country are classified by composition as follows: (1) cellulosic, composed of regenerated cellulose, cellulose diacetate, and cellulose triacetate; (2) synthetic, composed of polyamide, polyester, polyacrylic, polyvinyl, and polyolefin resins; and (3) inorganic, composed of glass and metal.

Numerous natural sources other than cellulose have been tried and usually abandoned. These include alginate from seaweed and regenerated protein from corn, peanuts, milk, soybeans, feathers, eggs, and even pumpkin seeds. The modern search turns toward the petroleum, natural gas, coke, acetylene, and related chemical fields. From these prolific sources come the monomers, or starting units, for addition polymerization (polyvinyls, acrylics, dinitriles, and olefins such as polyethylene and polytetrafluoroethylene), and for condensation polymerization (polyamides, polyesters, and polyurethanes).

The search for new fibers to add to the list already available is prompted by many factors. One reason is that the characteristic features required for the best performance in one specific use may be different from those of another application. By vastly increasing the variety of materials available to the textile industry, man-made fibers have revolutionized the design of fabrics. From an art concerned mainly with the creation of beautiful patterns, textile design has evolved into a branch of engineering by which improved levels of both aesthetic and functional quality can be achieved in clothing and home furnishings such as blankets, drapery, upholstery, and carpets.

Regenerated cellulosic fibers. To spin these fibers from cotton linters or wood pulp, it is necessary to put the cellulose into fluid form without destroying its chainlike molecular structure. About 1885, Sir Joseph Swan in England and Count Hilaire de Chardonnet in France conducted experiments leading to a process for extruding solutions of nitrocellulose and denitrating the filaments. In 1892, C. F. Cross and E. J. Bevan developed a better method using a solution of cellulose xanthate made from alkali cellulose and carbon disulfide. This is the basis of the viscose process used for the bulk of the rayon produced today by more than 6 companies in the United States. Some is made by the regeneration of cuprammonium cellulose (American Bemberg rayon) and some by the deacetylation of cellulose acetate (for example, Fortisan, manufactured by Celanese).

The viscose process consists of the following steps: steeping cellulose sheets in sodium hydroxide, shredding the sheets, aging the shredded material to form alkali cellulose, treating it with carbon disulfide to form cellulose xanthate, ripening and filtering the spinning solution, extruding the solution through the small holes of a spinneret into an acid coagulating bath, stretching the filaments, then washing, drying, and packaging.

From its humble beginning at the start of the century as artificial silk, rayon has steadily improved in quality, versatility, and economic importance. Partially responsible for expansion was the introduction of staple fiber which could be processed on the same types of equipment used for cotton and wool yarn textile spinning. For staple fiber, the number of holes in one spinneret is increased from the usual number (20–200) to as many as 20,000. Bundles of fibers from several spinnerets are combined into a rope or tow which can be cut into short lengths and packaged in bulk.

The second significant development was the attainment of high-strength yarns which led to cords suitable for automobile and truck tires, and to textile yarns with sufficient wet strength for launderable fabrics. The strength of fibers depends on many complex factors, including the arrangement of the molecules and the forces between them. In solution prior to spinning, cellulosic molecules exist in a random tangle of kinked and coiled chains. Stretching during spinning can straighten out the chains and orient them along the fiber axis to form a stronger arrangement. Because it is ineffective to stretch the uncoagulated spinning solution, the frictional drag in the spinning bath and the rate of coagulation must be adjusted properly.

Cellulose diacetate and triacetate fibers. Three *hydroxyl groups on each glucose unit of the cellulose* molecule can be esterified by acids such as formic, acetic, propionic, and butyric. The resulting cellulose derivatives can be spun into fibers. Of these, cellulose acetate is most important.

To make cellulose diacetate, the triacetate is formed first by treating cellulose with acetic anhydride and then converted back to the diacetate by partial hydrolysis, precipitated with water, washed acid-free, dried, dissolved in a solvent, usually acetone, and extruded by a dry-spinning process.

Cellulose triacetate (no hydroxyl groups) is less sensitive to hot water than the diacetate (one hydroxyl group) and garments made of it wrinkle less than those of regular acetate during laundering. It is sold under the trademark, Arnel.

Polyamide fibers. These fibers are made from polymers based on the reaction of a dibasic acid with a diamine or on the reaction of an amino acid with itself. Nylon is the generic term—not a trademark—for any long-chain synthetic polyamide used for fibers and plastics. The major one is 66 nylon made from a six-carbon diamine (hexamethylenediamine) and a six-carbon acid (adipic acid). Because the acid is dibasic, both ends react with amine groups to form amide linkages, thereby causing alternate units of the two starting materials to hook together in long chains. This type of reaction is called polycondensation, or condensation polymerization.

In 6 nylon, caprolactam molecules join together by self-condensation to form polycaproamide.

Although nylon can be dissolved in phenol or in formic acid the use of such unattractive solvents for spinning has not been necessary. Unlike many earlier fiber-forming materials, nylon can be melted without decomposition and pumped while hot through spinnerets. In melt spinning, the filaments congeal by cooling.

Nylon fibers, first produced experimentally by Du Pont in 1935, proved to be remarkable in many ways. After spinning, nylon can be cold drawn to about four times its original length. The accompanying reduction in diameter is permanent and occurs sharply over a short necking down zone which appears to move into the undrawn portion. In addition to producing a finer denier, the drawing process greatly increases the elastic modulus and the strength The alignment or orientation of the polymer chains along the fiber axis by either hot or cold drawing is now an important and rather typical process step for almost all synthetic fibers.

Another unusual property of nylon fibers compared to earlier man-made or natural ones is the ability to be heat set Heat setting was first used in nylon hosiery to prevent wrinkling during dyeing It is now used with fabrics of many types to achieve shape retention, permanence of pleats or ironed-in creases and to minimize wrinkling during laundering The basic mechanism is an adjustment of strong intermolecular forces by changes in the crystallinity. This has become an important part of the technology of other synthetic fibers.

In addition to its success in staple and filament fabrics ranging in weight from sheer hosiery to carpets, nylon has become the standard of excellent performance in rope, tire cord, and numerous industrial applications. It is made by an increasing number of fiber producers and is second only to viscose rayon in total poundage for man-made fibers. *See* POLYAMIDE RESIN.

Polyester fibers. This type of fiber is made from polycondensation polymers based on the reaction of a dibasic acid with a glycol or on the self-condensation of a hydroxy acid. The polyester fiber sold under the trademark Dacron is melt-spun from polyethylene terephthalate, the reaction product of ethylene glycol and terephthalic acid. Polyethylene terephthalate has excellent fiber-forming characteristics, high melting point, and hydrolytic stability.

The manufacturing steps as well as many of the uses for polyester fibers follow the general pattern established by nylon. In garments that require a high degree of resilience similar to that of dry wool plus an insensitivity to the plasticizing effect of moisture, polyester fibers are much superior to less hydrophobic fibers.

The commercial production of Dacron polyester fiber was started in 1953 in the United States. Terylene polyester fiber is manufactured in England and Canada. In addition, several companies have introduced pilot-plant quantities of polyester fibers based on other monomers. These include Kodel and Vycron. *See* POLYESTER RESINS.

Polyacrylic, polyvinyl, and polyolefin fibers. These are made from the addition polymers of unsaturated compounds. Examples of several types produced in the United States are shown in Table 1. Many others are made in Europe. Addition polymers are made from small molecules (monomers)

be ope

peatin[illegible]

copolymers can be made by using two or more different monomers in various ratios. The terminology nitrile alloy fiber has been applied to one spun from a mixture of two different polymers, for example, Zefran.

The double bonds responsible for linkage are no longer present after polymerization; thus polyethylene is a chain of methylene (CH_2) groups In acrylonitrile, one hydrogen on alternate CH_2 groups is replaced by a nitrile (CN) side-group Certain kinds of side-groups give rise through hydrogen bonding to rather strong interchain forces and are responsible for many useful properties. An example is the high melting point characteristic of polymers containing high percentages of nitrile groups. Many of the other vinyl and olefin fibers such as those based predominantly on polyvinyl chloride or polyethylene lack these attractive forces. Accordingly they generally are too heat sensitive for many textile uses and are limited to applications in which elevated temperatures are not encountered, or to a special application in which thermal shrinkage or heat bonding can be used to advantage. In contrast, a fiber made of polytetrafluoroethylene has an extraordinarily high melting point. It can be used for industrial purposes at temperatures up to 400–500°F.

Polyvinyl alcohol fibers are produced in Japan and Europe and are being considered for production in the United States.

Although the polymer of acrylonitrile had been known for many years, its successful use in fiber form depended upon the discovery of appropriate solvents for dry or wet spinning. It was also necessary to find satisfactory methods to dye the fiber The use of minor amounts of dye-accepting molecules introduced by the copolymer route and the development of new dyes and new dyeing tech

Table 1. Polyacrylic, polyvinyl, and polyolefin fibers

Trademark	Manufacturer	Type
Acrilan	Chemstrand	acrylic
Creslan	Am. Cyanamid	acrylic
Orlon	Du Pont	acrylic
Zefran	Dow	acrylic alloy
Vinyon HH	Am. Viscose	mainly vinyl chloride
Dynel	Union Carbide	vinyl chloride with acrylonitrile
Verel	Eastman	modified acrylic
Velon (Saran)	Firestone and others	mainly vinylidene chloride
Darvan	Goodrich	dinitrile (vinylidene nitrile)
Teflon	Du Pont	tetrafluoroethylene
Reevon	Reeves Bros	ethylene, propylene

niques has solved that problem. The first commercial acrylic fiber, Orlon, was introduced in 1948. The list of producers has steadily grown because of the attractive properties and versatility of fibers of this type. *See* POLYACRYLATE RESIN; POLYOLEFIN RESINS; VINYL RESIN.

Inorganic fibers. Glass and metals are among the *inorganic materials used for man-made fibers. The* increasing need in industrial and military fields for fibers which will stand extremely high temperatures has led to the use of quartz, aluminum oxide, and similar refractory materials.

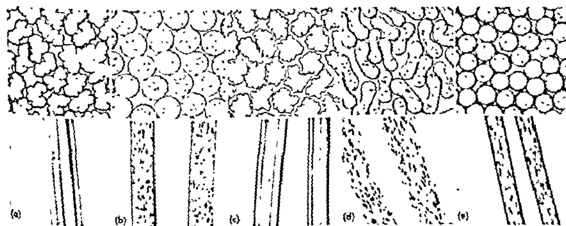
Metallic yarns have long been employed to add a decorative metallic glint and color to fabrics used for evening gowns, blouses, sweaters, tapestry, and upholstery. One modern method of preparation uses vacuum-plating techniques to apply aluminum coatings on strong, thin plastic films, such as Mylar polyester film. These films are then dyed and cut into widths ranging from 1/16 to 1/128 in. Such yarns are sold under a variety of trademarks including Reymet, Metlon, and Lurex. *See* GLASS AND GLASS PRODUCTS.

Specialty fibers. For many years natural rubber has been cut into thin ribbons and covered with a layer of cotton or rayon to form an elastic yarn used for items such as women's foundation garments and surgical hose. Some monofilaments are made by an extrusion process using either natural or synthetic rubber. In 1958, an experimental elastomer textile fiber called Fiber K was introduced. It can be spun as very fine yarns for weaving exceptionally light-weight elastic garments.

Examples of other special features incorporated in man-made fibers to achieve special effects are thick-and-thin yarns made by various methods such as a pulsating flow to the spinneret; colored fibers containing pigments added before spinning; mixed denier staple containing a range of filament diameters; hollow filaments; nonround cross section shapes of various geometric forms, such as crosses and triangles; and a variety of continuous filament yarns in which the fibers are crimped or looped in such a way as to give bulk and texture to the resulting fabrics.

Fiber properties. These are the physical, mechanical, chemical, biological, and geometrical characteristics of fibers. Some of the more important ones are tensile strength, elongation at break, *modulus of elasticity or stiffness, fatigue under re-*peated stress, resilience or ability to recover from deformation, moisture absorption and wettability, electrostatic properties, friction, color, luster, density, resistance to light, heat, weathering, abrasion, laundering, mildew, insects, chemicals, and solvents; and finally a number of geometric features such as diameter, cross-sectional shape, and crimp. Such properties play an important part in determining whether or not the fiber can be made into a fabric that will be wrinkle-resistant, pleasing to the touch, comfortable, easy to clean, durable, and attractive in color, luster, drape, and general appearance. With a knowledge of the physical properties of the fibers available to him, the textile engineer can choose the best fiber or best blend of several fibers to fit the intended use. The final result, however, is also dependent upon the proper choice and control of additional factors such as the yarn and fabric structure, the weave pattern, and the finishing of the cloth (*see* TEXTILE).

In spite of the research efforts put into the development of improved fibers, no fiber is perfect *for all uses. In many cases, perfect performance* represents an incompatible and therefore impossible combination, exemplified by the desire for sheer hosiery with heavy-duty wear. Notable among the deficiencies that still exist to varying degrees with some fibers both old and new are poor durability, soiling, static electricity, hole melting, yellowing with age, and poor shape retention. However, for each of these troubles a man-made fiber can be found which will solve the problem, as indicated in Table 2, which gives a partial listing of some of the major attributes of man-made fibers.



Cross-sectional and longitudinal views of typical man-made fibers. (a) Acetate. (b) Nylon. (c) Rayon. (d) Orlon. (e) Dacron. (*E. I. Du Pont de Nemours Co.*)

Table 2. Characteristic attributes of man-made fibers

Attribute	Properties required	Fiber recommended
Durability and abrasion resistance	High tenacity, high elongation at break	Nylon and polyester fibers
Resistance to soil particles	Smooth, round cross-section shape	Melt-spun fibers
Quick drying	Low moisture content inside fiber	Hydrophobic fibers: polyesters, polyacrylics, polyvinyls, nylon, triacetate
Winter warmth	Ability to entrap air in spaces between resilient fibers; thick fabrics	Most all synthetic fibers, the acrylics in particular
Summer comfort	Ability to wick perspiration from wet skin and dry quickly; thin but not too smooth fabrics	Polyacrylics and most rayon-synthetic blends
High resistance to light, solvents, acids, and bases	Compact chemical structure	Fluorocarbons and polyacrylics
Wrinkle resistance and wash wear	High resilience both wet and dry	Polyesters outstanding; many others good, including polyacrylics, nylon and triacetate
Freedom from static and hole melting	High moisture content	Rayon and rayon-synthetic blends
Crease and pleat retention	Heat settability	Nylon and many other synthetic fibers

All of the properties covered in Table 2 deal with functional performance. An additional category of importance is a group of the aesthetic properties of fibers and fabrics. By this is meant certain subtle qualities apparent to the hand, eye, and even to the ear. Examples of the latter are the so-called whisper of silk or the rustle of a good taffeta. One of the reasons that cellulose acetate was able to compete successfully with its predecessor, viscose rayon, was its superior handle—a soft, luxurious feel—especially desirable in crepes and satins. As the result of much technical effort, man-made fibers have been able to achieve those aesthetic qualities referred to by textile experts in describing the characteristic handle of a cashmere sweater, a wool flannel, or a silk taffeta fabric. See FIBER, NATURAL; POLYMER. [R.M.HO.]

Bibliography: American Chemical Society, *Literature Resources for Chemical Process Industries*, Advances in Chemistry Series 10, 1954; Harris Research Laboratories, Inc., *Handbook of Textile Fibers*, 1954.

Fiber, natural

A textile fiber of plant or animal origin. Science and technology have only quite recently lifted the natural fibers out of the realms of age-old industries based on hundreds of years of trial and error. Under the pressure of the highly competitive market situation and close economy of the industrial nations, some of the natural fibers were unable to withstand the fierce competition. One authority gave wool as providing 78% of all textile fibers, flax 18%, and cotton only 4% in the eighteenth century. By the end of the nineteenth century, cotton had usurped the place of wool as the primary fiber. Midway through the twentieth century, apparel fibers were distributed as follows: cotton, 71.6%; rayon, 17.3%; wool, 10.0%; synthetic and other man-made fibers, 0.9%; and silk, 0.2%.

This rise and fall in the history of natural fibers is also going on in the history of man-made fibers, which may be traced largely to the success of textile technology in a highly competitive market and its sensitivity to local economic conditions. As technology based on scientific research advances, man will be able to control the power of nature to an ever increasing extent. Already his mastery over the molecule is becoming more and more apparent in the textile industry. Although synthetic fibers can be made to demonstrate desirable properties under different conditions, vegetable and animal fibers are at the *mercy* of uncontrollable conditions, such as weather and genetics. For example, cotton strains must be kept within close tolerance by breeding; otherwise, quality will fall. The universal availability of natural fibers and the head production start, with their practical applicability, give these fibers wide use and commercial dominance. Natural fibers can be of mineral, vegetable, or animal origin. Except for asbestos, the mineral source is of very little commercial consequence (see ASBESTOS). On the other hand, vegetable and animal sources have great importance; and of these cotton surpasses all others. The animal fiber counterpart of cotton is wool.

VEGETABLE FIBERS

Cotton. The cotton plant is a member of the mallow family, Malvaceae, and because of its economic importance has received intensive study from both morphologists and physiologists. The genus *Gossypium* includes herbs and shrubs which attain heights of 10 in. to 6 ft. Although the plant is a perennial in original tropical habitat, it dies off with the first winter frosts and requires annual planting in temperate regions. In only a few countries, such as India, does it reach treelike proportions.

The flowering time is very short. After pollination, the cotton boll takes 50 days to ripen. Inside it, the ovules mature simultaneously with the covering seed hairs which are the source of the plant's vital importance to mankind, both as a fiber and as a cellulose source.

During the early stages of growth after pollination the hair grows only in a longitudinal direction, becoming tubelike and elongated, and it is surrounded by a wall of cellulose. The outside of the latter membrane consists of wax and fatty sub-

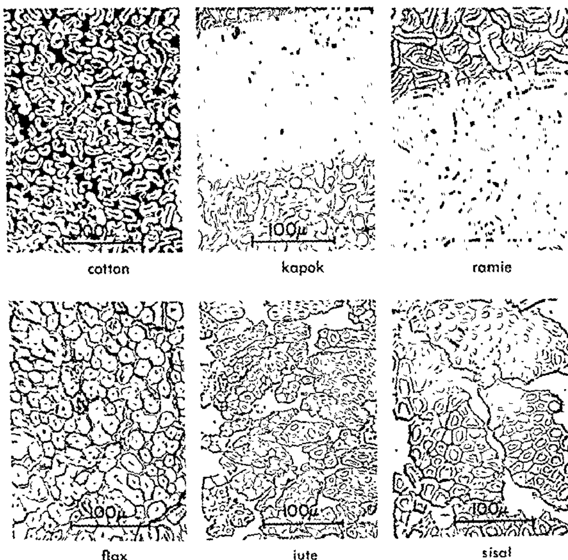


Fig 1. Important vegetable fibers. *(Institute of Textile Technology)*

stances, which are impermeable to gas and water and give the fiber its silkiness and luster After 24 days, another layer of cellulose is deposited on the inside of the cell wall. Concentric rings are formed, corresponding to the days of growth. The lint hair lives until the time of boll opening. As the lint hair dries, its tubelike form collapses and becomes very twisted in its length, exceeding the diameter from 1000 to 3000 times. The characteristic twists correlate with spiral fibril structures in the secondary wall. There are reversals of this twist during the convolutory growing process (Fig. 1).

Fiber. The grading of cotton fibers is based on various considerations, but the length of fiber is the most important. In the United States, because there are not many average staple variations in local areas, the trash (unwanted material) content is a primary consideration.

The most important universal standard grades are determined mainly by staple length. Sea Island cotton extends 2 in. in staple length, the finest Egyptian 1¾ in., long-staple American Upland 1½ *in.*, short-staple Upland about 1 in., Indian varieties ⅞–⅝ in., and Chinese ¾ in.

Although cotton thrives best in tropical and subtropical regions, there are places in southern Argentina and in Russia which are exceptions. The growth periods must be free of frost, and the normal period is 200 days in the cotton belt of the United States

Most of the cotton in the United States is now harvested mechanically. India, the second largest producer, has very little mechanization, whereas the Soviet Union, like the United States, is far advanced. The shortness of cotton fiber as compared to wool and flax fibers (which average two or three times longer) offers special problems to both the spinner and weaver. In addition, the natural curliness of wool assists spinning in contrast to cotton which has a finer diameter (10–20 μ) and which is

comparatively smooth except for the spiral twist of its convolutionary form and lateral folding. The specific gravity of cotton is 1.54. The tenacity varies from 2 to 10 g denier. This tenacity, unlike that of most other fibers, increases considerably when the cotton is wet; that of American Upland varies from 3.0 to 4.9, dry, and from 3.3 to 6.37, wet, measured in grams per denier. Its elastic recovery is poor as compared to that of the other fibers. American Upland's elastic recovery at 2% stretch is only 74, and 45 at 5%, whereas that of wool is 99 and 63, respectively.

Yarn. Yarn quality depends upon the origin, average staple length distribution, color, and feel of the cotton. In order to ensure a desired uniformity and product quality, batches of raw fiber, different in staple length and other characteristics, are processed by blending together. One important consideration is the detection of large proportions of immature and dead fibers because this variance will affect the spinning operation. It has been found that cotton can be overginned. This results in fiber damage through severe mechanical handling and desiccation by the air driers. Gin drying is carried out prior to removal of cotton fibers from the seed by gin saws. With the advent of machine-picked cotton, the ginner has had to contend with more trash *because the machines pick parts of stalk and leaves* and other components of the plant as well as the *bolls*. To overcome this, longer drying and more *elaborate* equipment are being used to clean the cotton. This tends to shorten the fibers and make them *more brittle* and compressionally stiffer.

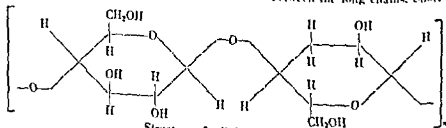
Although the ginner removes *much of the trash* through hot-air drying, often trash still remains *to* be removed later from the baled cotton by the *processer using opening* machines. The cotton fibers must then be parallelized *by carding to form* them into fine webs *which are drawn into slivers*. Slivers are again mixed with *other slivers. Drawing* the slivers with a slight twist on drawing *frames* containing multirollers produces then a greater degree of parallel *orientation. High-speed* drawing yields higher production rates, *but less-uniform* yarn. The sliver is then *twisted into roving* to give a more compact and stronger product. This *product* is wound into bobbins and then spun into yarn.

After the weaving operation, *the resulting fabric* may receive further processing *which can also be* complex, especially because modern textile technology can change cotton fabric more than any other material. Bleaching, *dyeing, and* printing can be considered among these, and lately with the introduction of resins, cotton can be given many types of properties, such as crease resistance, water repellence, and flame retardance. Although mechanical processes are involved, chemical technology is the promoter of these achievements.

Cellulose properties. The backbone of the chemical structure of cotton is cellulose, which may be regarded as a series of building blocks of glucose units (pyranose rings forming long chains), each with two primary and two secondary OH groups. The latter control the chemical reactivity of the whole structure.

The structural formula of cellulose shows how the individual anhydroglucose (pyranose) units forming the repeating units are linked at the one and four position through the β-glucoside bonds. The disposition of these chains in relation to each other determines the macroscopic properties of the fibers, that is, whether they are crystalline or unordered (amorphous). These two types of structural characteristic of cotton influence both its chemical and *physical properties*. The crystalline portions give the fiber strength, whereas the amorphous areas influence most of its rheological properties. Long-range elasticity is dependent upon amorphous structure which plays an active role in crease resistance. Cotton's cellulose is mostly crystalline, and the cohesion of the molecules is the result of van der Waals forces and fairly strong hydrogen bonding repeated many times between the glucose units without any primary chemical bonds, so that when the fiber reacts to a large strain consistent with fabric creasing, there are no direct *chemical* bonds to *promote the strained fiber* to its original configuration. Its reaction to such subjugation relies upon the entangled network of the amorphous molecular system. This accounts for cotton's poor elastic recovery. However, this deficiency is counterbalanced when the fiber is wet. The amorphous areas swollen with water molecules can act as an internal lubricant and release local strains in the long chain held in the more strongly bound crystalline regions, making for a more uniform stress pattern. Creases are thereby released, and the fiber is also allowed to carry heavier loads as a result of chain bonding by the water molecules.

The amorphous areas are chemically more accessible than the crystalline to reactivity. All the swelling takes place in the former areas, which are hydrophilic. Complete solubilization by water is prevented, however, by the strong hydrogen-bonding effects between the long chains. Short-chained cel



Structure of cellulose molecule

lulose such as is found in paper can be solubilized in water. The OH groups are capable of chemical addition such as esterification. With acetic acid, acetate rayon is formed, and with nitric acid, guncotton. Blocking of the OH groups reduces cotton's hydrophilic properties with complete chain retention, however. These reactive OH groups permit the use of chemical additives, for example, for flame resistance, crease resistance, and many other properties to increase the fiber's usefulness.

Although knowledge of its structure permits explanation of many of the reactions of cotton, others remain incomprehensible because of the randomness of the cellulose state. Cellulose exists neither in the completely crystalline state of pure glucose nor in the completely amorphous state of a supercooled liquid, such as is found in glass.

Bleaching. One of the most important wet-finishing processes is the bleaching operation. Warp sizing and impurities such as oil and dirt are removed by soap and surface-active agents. The fabric is then boiled in a kier, or vat, in which the noncellulose constituents of the fiber are removed by alkaline solutions. After the fabric has been boiled for several hours, the fats, proteins, and pectins are dissolved. These processes are known as scouring.

Severe treatment with alkali is able to break the glucosidic linkages. This reaction is known as oxidation degradation. Oxidation commences prefer-

These oxidative breakdown products are known as oxycelluloses. Because this heterogeneous reaction is complex, composition products are not well defined. Alkaline media are carboxyl-directive; that is, the pyranose becomes carboxylated through the 2,3-alcohol position.

After scouring, the goods are passed through the bleaching liquor to destroy all natural coloring matter in the cotton and to give the fiber a pure white appearance. In Europe and England, hypochlorite solutions are used for bleaching. In the United States, however, hydrogen peroxide is used in a continuous process. During the bleaching process, strength loss varies from 5 to 15%, and the degree of polymerization (DP) decreases from 2000 to 800. The DP is usually determined by the measurements of viscosity of solutions in cuprammonium or in cupriethylenediamine, which have no influence on the glucosidic bonds.

When cotton is damaged, the DP will fall as low as 500 or even lower. Acid damage hydrolyzes the glucose chain to form the hydrocellulose and can also result in the formation of carboxyl and carbonyl groups from the hydroxyl groups. This increase of reduction products can be detected suitably by precipitations of silver or copper complexes, or by means of such basic dyestuffs as methylene blue. After bleaching, mercerizing can be undertaken when, under suitable tension, the fibers are immersed in low concentrations (about 20%) of caustic. The resulting swollen fibers take on a pleasing luster. The cellulose's crystallitic structure is improved by increased orientation, thereby giving added strength. The fibers are also straightened in the operation by removal of the convolutions. See BLEACHING.

Dyeing. With the invention and development of synthetic dyestuffs, the process of dyeing cotton both in the yarn and fabric has become a technological triumph of rapidity, even coloring, and control. Carefully skilled pretreatment operations, optimum choice of auxiliaries for preparing the material for the dyeing, and judicious selection of the best color for end-use considerations such as sunlight and washing fastness are some of the factors responsible for this success. Cotton dyes show a remarkable affinity for the fibers. Chemical brighteners used by commercial laundries have also enhanced this success. This is not true of all other cellulosic fibers. There is no actual chemical linkage between the dye molecule and the cellulose. Cotton is readily receptive both to surface-active agents with its unique fine structure and its swelling characteristics as well as to the entry and adherence of dye molecules which become permanently fixed. Some new types of dye are claimed to have chemical bonding which forms strong ester or ether linkages with the cellulose, with brilliant fast colors resulting.

Cotton has found its widest application in creation of shade-style demands, not only because of the ease with which it may be colored by dyeing and printing, but to an even greater extent because of its ability to undergo the chemical reactions of oxidation and reduction in the fiber substrate during the color fixing and color discharging.

Cotton can be considered, in comparison to other fibers both old and new, to be highly versatile, and it lends itself to processibility necessary to meet most textile requirements.

Although new man-made fibers are competing increasingly for some of the markets created by cotton, cotton is still assured a leading position as a cheap, all-purpose fiber for some time to come.

Cotton, the most universal fiber, is one of the major sources of cellulose. This is because it is in world-wide and plentiful supply, and because its cellulose content ranges from 88 to 96%. See CELLULOSE; COTTON; DYEING.

Bast fibers. This special group includes all fibers stripped from the inner bark of plants such as flax, hemp, jute, ramie, and others of lesser commercial importance. They are used extensively in the textile and paper fields.

Flax. Flax is the most important in the group because it is widely used as a textile fiber. Russia is the greatest producer of linen fiber, but some of the best linen comes from Belgium and Ireland. The flax plant is not robust and demands rich soil, careful attention, and cool, damp climate.

The flax fiber has a high percentage of cellulose (almost 90%), and high crystallinity and molecular orientation, both of which are greater than those of cotton. This accounts for its having a strength

greater than that of cotton, but a lower breaking extensibility. The prevalent crystallite structure acting like spot welding permits little chain slippage. The linen fiber's natural color varies from pale cream to gray, depending upon the extraction process of retting. Bleached, good-quality linen is pure white.

Although pure flax and cotton have very similar chemical composition and properties, the gross fiber structure is different; the linen fibers are composed of a fibrillar structure containing innumerable fibrils together in a cemented network of intercellular tissue composed of pectin and waxy material (Fig. 1).

Linen is more heat-sensitive than cotton; boiling water will injure the fibers, and care must be taken in ironing and pressing. Linen, unlike cotton, dyes poorly and has poor dye fastness. It is also much more easily harmed by bleaching, which must be carried out very carefully to prevent fibrillation and degradation.

The high cost of linen is not only in the growing and harvesting, but also in the cost of the separation of the fibers from the flax stalk, which is essential before processing into yarn. A process called retting is best carried out slowly in tanks in order to give the most superior textile; however, more-rapid chemical methods are being developed. After retting, which is a degumming process, the fibers must be mechanically loosened and pried by rupturing the so-called fiber shive bond. Various machines using the motions of shearing, crimping, scraping, beating, and oscillation are still in use today. Refinement of linen fiber has not yet become proficient enough to allow total processing in cotton machinery. Low percentages of cotton have to be blended into the linen, although research attempts have been made to overcome this. *See* TEXTILE MICROBIOLOGY.

Throughout history flax has consistently attracted mankind. Besides its great durability, as exemplified by the cloth covering the Egyptian mummies which was unearthed in a good state of preservation, it has a superior heat conductivity and feels cool. The hygroscopic properties of linen give it one of its most desirable qualities. Initially, moisture will spread rapidly through the interstices of the fabric at a rate faster than in cotton; at the same time, it will not penetrate the fiber with much absorption. This process [illegible]

Jute. This is the cheapest of natural textile fibers. It is also used for packing material and as draperies and wall coverings. It readily takes direct dyes. Most of the jute comes from the Indian Peninsula, with most of the harvesting done in India and the processing in Pakistan. Work is underway to improve the fiber properties by breeding and more efficient bleaching (Fig. 1). Acetylation of the fiber may also throw open wider markets for textile use. Its outstanding quality is, however, its truly remarkable resistance to microbiological organisms. *See* JUTE.

Ramie. Another textile fiber used by man for thousands of years, especially in the Orient, ramie also has a strong microorganic resistance, and is little impaired by water. It has tremendous tensile strength, resulting, as in flax, from high molecular crystallinity and orientation (Fig. 1). Some of the fire hoses of England during the bombing of London in World War II were constructed of ramie. They withstood all the harsh treatment successfully, both from high pressure and impact. The fiber has a high, pleasing luster. *See* RAMIE.

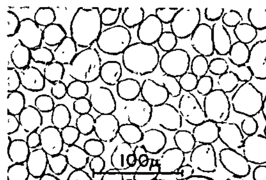
Hard fibers. This is another distinct class of vegetable fibers, derived mostly from the leaves of tropical and subtropical plants. Manila hemp is one of the most satisfactory of these, used in rope making in the amount of 520,000 tons annually. Since 1946, hard fiber production has increased. Depending upon the plant source, the fibers can be put to a variety of uses, ranging from the sheerest of fabrics, such as veils, to heavy ropes. The major portion of the fiber composition is cellulose, the minor, hemicellulose. *See* HEMP.

Coir has commercial importance and is mostly used for matting and packing. Almost all the world supply comes from Ceylon and India. The source of the fibers is coconut husks. The highest grade coir is removed prior to nut shipment and copra extraction. The coir is retted, after separation, in a manner comparable to flax retting. The process may require several days, months, or longer. The fibers are thoroughly and repeatedly cleaned, then carded and dried. Much of the cleaning treatment is now carried on with machine equipment. Compared to other fibers, its tensile strength is low; however, it has relatively high elasticity and is highly resistant to rotting, especially in sea water. This property has for centuries given the fiber use in cordage ropes, fishing nets, and other nautical items. *See* COIR.

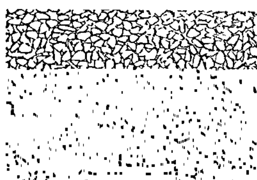
ANIMAL FIBERS

Wool. This fiber has also been used since ancient times. Sheep's wool is the most valuable of animal hair, excelled only by that of the vicuna, cashmere goat, angora goat, and camel. *See* CASHMERE; VICUNA.

One of the most famous and sought-after wools is that of a mountain sheep bred in Spain, the merino. Both environmental conditions and breeding influence fiber quality. The lowland sheep have coarser wool with less crimp; various breeds are conveniently classified, therefore, into mountain and lowland types. China and Mongolia supply most of the camel wool, and most cashmere comes from India and Tibet. Wool production in America and Europe is steadily declining because of unfavorable conditions. On the other hand, sheep farming in Australia, South America, and South Africa,



wool



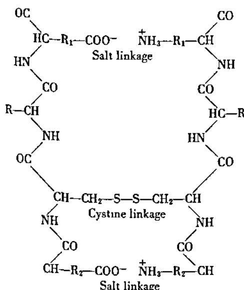
silk

Fig. 2. Important animal fibers. (*Institute of Textile Technology*)

areas with more favorable conditions and wide-open spaces, are able to export high-grade wool.

No synthetic fiber can match wool's quality characteristics, especially its insulating qualities. However, with the advent of man-made fibers, its weaknesses have become more evident, and wool research has taken on an international aspect in its endeavor to maintain its position among world-wide commodity fibers.

Modern machinery, equipment, and methods are assisting an industry which has been technologically backward. Electrically driven shears, making for close and rapid cropping, remove the fleece in one piece. An operator can clip up to 90 sheep per day. Shearing is usually carried out biannually. At the same time, modern dyeing and finishing techniques are helping the fiber to hold its own



R = groups linking other chains

R_1 = bulky amino or side-chain groups such as are found in proline and tyrosine

R_2 = short-chain linkage groups

WOOL CHEMICAL STRUCTURE

Wool structure. Quality, characterized by fiber fineness and other properties of the wool, depends upon the area of its bodily origin on the sheep. For this reason, it is sorted into uniform batches for spinning.

The chemistry of the wool fiber is that of a highly complex proteinaceous material. Histologically, the fiber structure is made of several layers of different types of keratin cells. There is an outer layer with three subdivisions, then the cortex with its cortical cells, which enclose the medulla or core. The core is often absent. The cortical cells frequently contain natural pigments. The outer layer, containing overlapping scales, gives wool its remarkable fiber surface appearance and partly its ability to felt. This scaly outer layer is composed of keratin, whereas keratin C, containing tyrosine, exists mostly in the fiber's interior (Fig. 2).

Keratin is the paramount substance of wool fiber and is similar to most proteins of α-amino acids which form chains (polypeptides) of varying composition and complexity, dependent upon the type of α-amino acid present. These long peptide chains are bridged by a cystine and salt linkages, with pairs of sulfur groups in dual attraction. Some amino acids have been used to exemplify the structure of wool Knowledge of the chemistry of wool, like that of cotton, still remains limited.

Properties. The chemical structure plays an important role in the physical behavior of the fiber such as the ability to recover from deforming forces when in fabric form. The cystine-bond formation may be broken under straining forces, but the amino-acid bonds are comparatively much stronger. Much weaker, however, are the salt linkages which are governed by van der Waals forces as in the hydrogen bonding of cotton cellulose. Because the side chains of the main polypeptide chain are so bulky, there is no close packing as in cellulose chains, and they tend to fold into coils. These coils can be stretched out at the expense of side-chain-bond breaking as the chains slip by each other, preserving fairly secure chemical stability. This permits the wool to have great extensibility, but at the same time new bonds can be formed and one

molecular form of keratin will change to another, thereby retaining its original strength before stretching. This molecular disposition provides for the superb garment properties—crush resistance, warmth, and general feel of comfort—of the wool.

Heat with pressure permits creases to be pressed or ironed into wool fabric, giving it thermoplastic properties, with the breaking of the salt linkages.

Although wool has a low strength of 1.0–1.7 g/denier, its elongation to rupture ranges from 35 to 50%, dry and wet, respectively, higher than that of any other natural fiber. Wool fabric is fundamentally a good insulator because it can retain a large amount of entrapped air during all types of fabric movement. It feels warm because of the fiber's roughness and natural crimp. Wool can absorb more water than any other fiber without feeling wet or clammy; hence, it has a high wicking resistance. Wool has fairly good processing properties, for dyeing, washing, scouring, and reactivity with certain dyestuffs This characteristic stems from its hydrophilic behavior and its ability to be swollen readily by various classes of dyestuffs. There are only a few exceptions to the large number of color combinations which are fast to wool. In addition. by virtue of the amphoteric alanine groups, acid, chrome, metallized, vat, and basic color can be used

Dyeing. Dyeing is one of the most important finishing operations in the woolen industry. It is carried out on yarn, piece, slubbing, and loose goods. The dye process involves some complex chemical reactions, both because of the complexity of the dye molecule and because of the fine structure of the wool. The extensive ranges of dyes available for wool permit the selection of a color for shade and fastness dependent upon the ultimate end use of the finished wool.

Processing. A variety of yarn and fabric process systems is available for the manufacture of the raw wool into cloth. The system used is determined by the end use of the finished material. Worsted yarns are made from long-staple fibers combed free of short fibers. They are smooth in appearance and may be spun to a high count. Woolen yarn often contains short fibers and reworked waste. These yarns are coarse and find uses in carpeting and rugs. Different types of knitting machine are also available for the many styles of fabric.

Wool is one of the few fibers which has the unique property of being able to felt. This is a desirable feature in milling and scouring and for felt making, but undesirable in all-wool styles, such as stockings, sweaters, swim suits, and other garments. This has been partly overcome by the chlorination process, which uses dilute solutions of chlorine and can be employed at most stages of manufacturing.

It is not fully understood why felting occurs, but it may be a result of the tension existing in the wool, the peculiar serrated shape of the outer layers of the fiber made by the pointed-tip overlapping scales, and the chemical characteristics of the cystine linkages. Wool is sensitive to alkali damage, but more resistant to acid than cotton or the vegetable fibers.

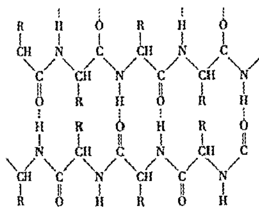
Storage of wool garments and carpets is problematic because of moth-larva and carpet-beetle damage, respectively. Suitable chemical pretreatments have overcome this defect to some extent. Suitable organic chemicals placed in the dye bath are now used, the effectiveness of which lasts for several years without detracting from the qualities of the fibers. Other agents can also be used in the cleaning fluids to give the garment equal substantivity and to be active against moth damage over a period of time. See WOOL.

Silk. The manufacture of silk can be traced in many of the old materially flourishing nations, including ancient China 5000 years ago. Silk was one of the richest of merchandise traded by all the leading commercial nations of the past and has been until modern times. Of all the fibers, silk is unsurpassed for its beauty, drape, comfort, appearance, and appeal.

This fine-quality product, however, is dependent upon the growth of the mulberry tree, and also upon the development of the silkworm, both of which are at the mercy of environmental factors. Also, silk can be produced only in low-cost labor areas because no process has been evolved to eliminate successfully the hand reeling of the cocoon strands. The use of silk declined steadily in the United States after 1920. New synthetics became popular, until by 1929 the import of silk bales fell to 662,000, of which 29% was used for weaving. By 1939, the bale consumption was reduced to only 385,000 bales, and 9% reached the weaving industry. By 1952, it had dwindled to about 40,000 bales During World War II silk was used exclusively for parachute and other defense uses, and none was given civilian outlets.

Properties. Silk is a strong fiber. It has a tensile strength of 4–6 g/denier and is exceeded only by two other natural fibers, linen and asbestos; glass, Dacron, nylon, and Fortisan are synthetic fibers of superior strength. Silk and nylon are the toughest fibers. Toughness is the amount of energy needed to rupture the fiber, based on unit mass. Silk combines strength with a high degree of elasticity to produce this remarkable toughness. Silk has excellent resistance to shearing forces, a factor in warmth retention of garments, and is far superior to nylon, excelled only by two other nonapparel fibers, glass and kapok. Silk fabric has several advantages, including the resisting of creases in apparel, especially after special finishing treatments. Silk gives the wearer a feeling of warmth in winter and coolness in summer because of its good insulation qualities (Fig. 2).

The fiber is degraded slowly by direct sunlight and tends to yellow with age. It is difficult to dye with fast colors, the brightest of which are the poorest in this respect.



Simplified chemical structure of silk

Structure. The chemical properties of silk can be explained by its structure. The long-chain molecule is made up mostly of extended glycyl-alanyl-serine chains. Existing between these chains are numerous secondary bonds made by hydrogen bridges, accounting for its high tensile strength. The extended condition of the chains and the high proportion of molecular orientation limit the extensibility found in the other polypeptide-chain fiber, wool. The presence of tyrosine, proline, and serine groups give the fiber strong polar tendencies. This dissimilarity to nylon structure is shown by fiber swelling although the same type of hydrogen bonding is present. The amorphous structures of the fiber made by the bulky organic radicals are unable to align in chains and permit access by reactive chemicals

Most of the silk supply is at present raised in Japan, where technological improvements are underway, such as the development of reeling machines. Stronger silk is being investigated by special types of breeding. Basic research is also in progress to produce silk adaptable for high-speed spinning and weaving equipment, by obtaining silk bred with a uniform average thickness and strength enough for mechanical reeling. Also, chemical research is underway to obtain a silk solution satisfactory for spinning comparative to the spinning dope used in synthetic-fiber spinning. *See* SILK.

See also BLEACHING; DYEING; FIBER, MAN-MADE; TEXTILE; TEXTILE CHEMISTRY, TEXTILE PRINTING. [C.C]

Bibliography: H. R. Mauesberger (ed.), *Mattheus Textile Fibers*, 6th ed., 1954; J. M. Preston (ed.), *Fibre Science*, 1953.

Fiber bundle

A type of structure arising frequently in various branches of mathematics, especially differential geometry, Lie groups, and algebraic geometry. A fiber bundle is a principal tool for the applications of algebraic topology to these subjects

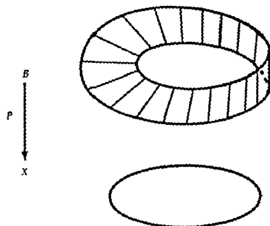
A fiber bundle is a topological space B which is decomposed in a smooth fashion into a family of closed disjoint subsets, called fibers, each of which is homeomorphic to a fixed space Y (*see* TOPOLOGY). The base space X of the bundle is the space whose points are the fibers. The projection $p:B \to X$ is the function assigning to each point the fiber which contains it, and X is topologized so that p is continuous. The requirement of smoothness reads as follows: for each $x \in X$, there is a neighborhood V of x, and a homeomorphism $\varphi:V \times Y \to p^{-1}(V)$ satisfying $p\varphi(x',y) = x'$ for all $x' \in V$ and $y \in Y$.

The product space $X \times Y$ and its projection into X satisfy these conditions with each $V = X$. The Mobius band (see figure) is one of the simplest examples of a bundle which is not a product. The fibers are the line segments perpendicular to its center line, and its base space is a circle. The corresponding product space is a cylindrical surface. Because of such examples, fiber bundles are sometimes called twisted products.

Fiber bundles arise naturally in several different ways. If B is a Lie group and Y is a closed subgroup of B, then the family of left cosets of Y in B is a fibration of B, and the base space X is a manifold on which B is a transitive transformation group. For example, let B be the group of all rotations of a 2-dimensional sphere, and let Y be the subgroup of rotations about a fixed axis. Then B is a 3-dimensional manifold (projective 3-space), Y is a circle, and X is the sphere.

Let X be a differentiable manifold. The vectors tangent to X at a point $x \in X$ from a linear space Y_x. Their union for all $x \in X$ is a differentiable manifold B called the tangent bundle of X. The projection p assigns to each vector its initial point. Similarly, there is a tensor bundle over X for each algebraic type of tensor (for example, skew-symmetric tensors of two covariant indices with nonzero determinant).

A cross section of a bundle B is a mapping $f:X \to B$ such that $pf(x) = x$ for each $x \in X$. In case $B = X \times Y$, f is the graph of a mapping $X \to Y$. In case X is a differentiable manifold and



Projection of topological space B (a Mobius band) its base space X (a circle).

B is a tensor bundle over X, then f is a tensor field.

Many bundles do not have cross sections. For example, the 2 dimensional sphere does not have a continuous field of nonzero tangent vectors. A method of algebraic topology called obstruction theory can be applied in many cases to decide whether a cross section exists. It leads to cohomology classes of the base space called characteristic classes of the bundle. Study of these has led to deep and far-reaching results in differential and algebraic geometry. For example, the 4-dimensional sphere does not admit a non-Riemannian metric of the Lorentz type. Another example: if $n \neq 2$ or 6, then the n-dimensional sphere does not admit a complex analytic structure.

The homology groups of B, X, and Y are closely related. They are imbedded in an algebraic system called the spectral sequence of the bundle. In special cases, the homology of B may be computed in terms of that of X and Y and of certain invariants which express the amount of twisting. In all cases, the Betti numbers of B never exceed the corresponding Betti numbers of $X \times Y$ The homotopy groups of B, X, and Y are more simply related They are the terms of an exact sequence

$$\cdots \rightarrow \pi_{q+1}(X) \rightarrow \pi_q(Y) \rightarrow \pi_q(B) \rightarrow \pi_q(X) \rightarrow \cdots$$

Because of these relations, the topological structure of a space may be analyzed by choosing a *suitable fibration*. For example, the real orthogonal group in n variables $O(n)$ is fibered by the left *cosets of* $O(n-1)$, and the base space is a sphere S^{n-1} of dimension $n-1$. When n is even, the twisting invariants are zero, and when n is odd, *they are zero modulo 2. By* an induction on n, it follows that the Betti numbers of $O(n)$, based on coefficients modulo 2, *are those* of the product space $S^0 \times S^1 \times \cdots \times S^{n-1}$. *See* MANIFOLD.

[N.E.S.]

[illegible]
M [illegible]
M [illegible]
de [illegible]
K. N [illegible]
Mat [illegible]
The [illegible]

Fiber crops

Plant fibers are long, thick-walled cells, or groups of cells, that form slender strands. They may be divided into fairly specific categories according to their location in the plant. *See* PLANT ANATOMY.

Long, multiple-celled fibers may be subdivided into hard or leaf fibers that traditionally are used for cordage, such as sisal for binder and baler twine, and abaca for ropes; soft or *bast* (stem) fibers used for textiles, for example, flax for linen and jute for [illegible] ... such as Spanish moss for upholstery; or fruits, such as coir used for cordage, floor coverings, sacks, and coarse textiles.

Short, one-celled fibers come from the seeds or seed pods of plants such as cotton, milkweed floss, and kapok. Uses include cloth, cordage, and insulation.

Wood fibers come mostly from the trunks of coniferous and other softwood trees, and are used mainly for paper and cellulose products (*see* CONIFERALES).

There are many exceptions to these uses. For example, some sisal near the sources of production is used for bags, and jute is used for cordage. However, the large quantities that enter international trade are used as indicated above. *See* ABACA; CELLULOSE; COIR; COTTON; FLAX; JUTE; KAPOK TREE; LINEN; PAPER AND PAPER PRODUCTS; SISAL; *see also* FIBER, NATURAL; FRUIT (BOTANY); LEAF (BOTANY); ROOT (BOTANY); SEED (BOTANY); STEM (BOTANY).

[E.G.N.]

Fibrinogen

A plasma protein probably synthesized in the liver, fibrinogen is the precursor of fibrin, the insoluble fibrous protein formed during blood coagulation. The concentration of fibrinogen in plasma increases in pregnancy and inflammatory states, increasing the rate at which red blood cells sediment in shed blood. Congenital or acquired deficiencies of circulating fibrinogen may be associated with severe disorders; transfusion of partially purified fibrinogen, which is concentrated in fraction I of human plasma by E. J. Cohn's method, may provide temporary relief. Fibrin forms in wounds; its function in healing is obscure. Material histochemically similar to fibrin is found in lesions of many inflammatory and hypersensitive states. In surgery, sheets of preformed fibrin foam mixed with thrombin are used to promote hemostasis. *See* BLOOD.

[O.D.R.]

Fidelity

The degree to which the output of a system accurately reproduces the essential characteristics of its input signal. Thus, high fidelity in a sound system means that the reproduced sound is virtually indistinguishable from that picked up by the microphones in the recording or broadcasting studio. Similarly, a television system has high fidelity when the picture seen on the screen of a receiver corresponds in essential respects to that picked up by the television camera. Fidelity is achieved by designing each part of a system to have minimum distortion, so that the waveform of the signal is unchanged as it travels through the system. *See* DISTORTION (ELECTRONIC CIRCUITS); SOUND REPRODUCTION SYSTEMS, ELECTRICAL.

[J.M.R.]

Field emission

The emission of electrons from the surface of a conductor into vacuum (or into an insulator) under influence of a strong electric field. In the field emission process, electrons penetrate through the sur-

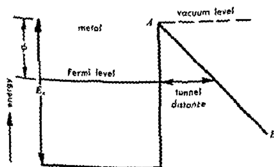
face potential barrier by virtue of the quantum-mechanical tunnel effect; this effect is a result of the wave nature of elementary particles and has no classical interpretation. Field emission differs from thermionic, photo-, and secondary emission in that in these processes the electrons escape across the surface potential barrier.

The process is often called cold emission, the adjective "cold" referring to the fact that it does not require elevated temperatures; the process has been observed at temperatures approaching absolute zero. However, field emission may be enhanced by heating because the probability for tunneling through the surface potential barrier increases with increasing energy of the electrons. For a diagram of the surface potential barrier, see SCHOTTKY EFFECT.

For a metal at absolute zero temperature the process may be discussed with reference to the illustration. The dashed "vacuum" level represents the energy of an electron at rest in free space; the energy of a conduction electron at rest in the metal lies below the vacuum level by an amount E_r. The energy levels below the Fermi level in the metal are occupied by conduction electrons, whereas those above it are empty. In the presence of a strong constant field F produced by a positive anode, the potential energy of an electron outside the metal is represented by the line AB. An electron in the metal has a reasonable probability to tunnel through the barrier if the tunnel distance is of the order of 10 angstroms or less. For a work function ϕ of, say, 4 electron volts, a field of the order of 4×10^7 volts/cm is thus required to produce an observable tunnel effect for electrons near the Fermi level. For additional details see FREE-ELECTRON THEORY OF METALS; WORK FUNCTION (ELECTRONIC). From the known energy distribution of the conduction electrons and for a given shape of the potential barrier, the emission current density J may be calculated. The field dependence of J is well represented by the Fowler-Norheim formula

$$J = BF^2e^{-\beta/F}$$

where B and β are constants which depend on the exact shape of the surface potential barrier.



Illustrating the energy level scheme for field emission from a metal at absolute zero temperature

Experimental verification of the theory requires accurate knowledge of F and hence extremely smooth surfaces; at high temperatures, surface migration of electrons on a clean metal point produces a perfectly rounded, nearly hemispherical tip. If a fluorescent screen at anode potential is placed opposite to the tip, a greatly enlarged picture of the emitting surface is obtained (field-emission or field-ion microscope, invented by E. W. Müller in 1937) which allows detailed studies of various surface effects. Current densities as high as 10^9 amp/cm^2 have been observed. Positive ion emission has been observed under influence of reverse fields of 10^8 volt/cm. See FIELD-EMISSION MICROSCOPY; see also ELECTRON EMISSION.

[A.J.DE.]

Field theory, classical

Classical field theory, more commonly known as continuum physics or continuum mechanics, is the study of distributions of energy and matter under circumstances where the discrete nature of the latter is unimportant. This will generally be the case for a system of a (usually extremely) large number of particles. In this category is the study of the flow of fluids, heat, and other forms of energy; electromagnetic currents and waves, including optical phenomena; and macroscopic theories of the elastic and plastic deformation of solids. Consequently, most of applied physical science, such as strength of materials, hydraulics, heat transfer, and aerodynamics, is included.

The quantities involved in these theories, for example, mass and energy densities, fluid velocities, and electric field strengths in a material medium, may be regarded as local averages of corresponding microscopic variables. In many cases, it has been possible to obtain the macroscopic field equations by an appropriate averaging of the microscopic motions over space and time intervals that are small macroscopically, but not microscopically. For example, see STATISTICAL MECHANICS.

Mathematically, field theories are characterized by equations having partial derivatives with respect to position and time. These equations may express the requirements of conservation of energy, or linear and angular momentum, and also the geometrical properties of the particular type of field.

So-called constitutive equations must be added to the general equations. These define ideal materials by restricting the field variables in some way. Simple examples are the assumptions of incompressible or nonviscous fluids, of perfect gases, or of magnetic materials without hysteresis.

Many of the mathematical difficulties of continuum physics stem from the nonlinearity of the equations. In order to obtain tractable linear equations, the assumption is often made that the system departs from a simple type of motion by only a small amount.

Gravitational and unified field theories, although not a part of quantum theory at the present time,

are not usually considered to be a part of continuum physics because of their ultimate concern with the structure of elementary particles. Quantum field theories arise from the dual wave-particle nature of matter. *See* GRAVITON; QUANTUM FIELD THEORY; UNIFIED FIELD THEORIES; *see also* THEORETICAL PHYSICS.

[B.C.]

Bibliography: W. V. Houston, *Principles of Mathematical Physics*, 2d ed., 1948; G. Joos, *Theoretical Physics*, 3d ed., 1958; L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, vols. 6 and 7, 1959; A. Sommerfeld, *Lectures on Theoretical Physics*, vol. 2, 1950, vol. 3, 1952.

Field-emission microscopy

A technique that uses field emission of electrons or positive ions to produce a magnified image of the emitter surface on a fluorescent screen (*see* FIELD EMISSION). The technique was first proposed by E. W. Muller in 1936. The field-emission (or field-ion) microscope is the most powerful microscope known. The tip of the needle-shaped emitter shown in Fig. 1 is projected directly without lenses, with a magnification up to 1,000,000 diameters and a resolution, with electrons as the emitting particles, of 20 A. Specimen tips made of 22 different pure metals and of several alloys and conducting compounds have been similarly magnified. With this microscope, adsorption of various gases has been studied in the temperature range from 4°K to above 2000°K.

The condensation and surface migration of evaporated materials and surface reactions important in catalysis have been investigated. Medium-sized individual molecules such as phthalocyanin have been made visible.

If the microscope is operated at liquid hydrogen temperature with the tip positive, and helium is introduced, helium ions depict the tip surface with a resolution up to 2.3 A. The field-ion microscope is capable of resolving the individual metal atoms on

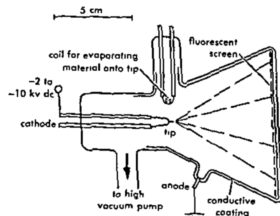


Fig. 1. Field-emission microscope operated with electrons.



Fig. 2. Atomic structure of a nearly perfect platinum crystal observed in a field-ion microscope.

the surface of a crystal lattice as shown in Fig. 2. Application is limited to about 10 metals having high melting points, but valuable information on lattice imperfections, such as dislocations due to stress or radiation damage, and on other structural details of atomic dimensions has been obtained.

[E.W.M.]

Bibliography: R. H. Good, Jr., and E. W. Müller, *Field Emission*, in S. Fluegge (ed.), *Handbuch der Physik*, vol. 21, 1956; E. W. Müller, *Field Ionization and Field Ion Microscopy*, in L. Martin (ed.), *Advances in Electronics*, vol. 13, 1960.

Field-enhanced emission

An increase in electron emission resulting from an electric field near the surface of the emitter; the field may be applied externally, or it may be produced as a result of the electron emission itself. The thermionic emission from a cathode in a vacuum tube may be enhanced by raising the voltage of the anode which collects the electrons; the electric field thus produced near the cathode reduces the effective thermionic work function.

The secondary electron emission from an insulating layer on a metallic back-plate may be enhanced as a result of the secondary emission process itself. If the secondary yield is larger than unity, the emitting surface will charge positively with respect to the back-plate, thus creating an electric field which aids the electrons in escaping from the surface. The fields so produced may be very large if the insulating layer is thin; for example, a potential difference of 1 volt across a layer of 1000 angstroms thickness creates a field of 10^5 volt/cm, which is of the same order as the dielectric breakdown field of insulators. *See* SCHOTTKY EFFECT; SECONDARY EMISSION; THERMIONIC EMISSION.

[A.J.DE.]

Fièvre boutonneuse

A generally mild febrile disease of man due to *Rickettsia conorii*, a bacterialike organism. While the disease resembles Rocky Mountain spotted fever in some respects (for example, the rash), it differs in exhibiting a characteristic tache noire, or primary ulcer, at site of tick attachment, swollen lymph glands which drain the area, and furthermore, it does not exhibit localities of high virulence. The generalized rash has a similar maculopapular appearance and includes the palms and soles, but the case fatality rate is less than 3%.

The disease occurs in parts of the Old World, particularly the Mediterranean littoral, the Crimea, and parts of India. The primary cycle in most areas involves domestic dogs and the brown dog tick, *Rhipicephalus sanguineus*, which invades premises inhabited by their hosts. Again, as in spotted fever, ticks are the probable reservoirs. *Rickettsia conorii*, the bacterialike causative agent, has characteristics similar to *R. rickettsii* in morphology, growth requirements, and animal reactions. However, spotted fever vaccine will not immunize guinea pigs to challenge or infection with these milder strains. Because of the sporadic incidence and low mortality, no vaccine has been developed or needed.

A strain isolated from *R. sangüineus* ticks from the premises of a former patient in Kashmir, India, where the disease is known as Indian tick typhus, behaved in the laboratory like Mediterranean fièvre boutonneuse. *See* RICKETTSIALES; RICKETTSIOSES; SPOTTED FEVER, ROCKY MOUNTAIN. [C.B.P.]



Fièvre boutonneuse showing rash and tache noire in popliteal area. (*From C. B. Philip, in R. L. Pullen, Communicable Diseases, Lea and Febiger, 1950*)



Rickettsia conorii, causative agent of fièvre boutonneuse, in stained smear of infected (Morocco strain) yolk sac of chicken embryo. (*From C. B. Philip in R. L. Pullen, Communicable Diseases, Lea and Febiger, 1950*)

Fig

Ficus carica, a deciduous tree with a milky juice growing in subtropical regions of Asia, Africa, Europe, and the United States. The fruit, known as a syconium, consists of a fleshy hollow receptacle lined with sterile male flowers and with a small opening at the end *See* FRUIT (BOTANY). Most edible types and varieties develop fruit without pollination (*see* REPRODUCTION, PLANT). Smyrna-type figs require pollination by a small wasp, *Blastophaga psenes*, which lives primarily in the inedible, pollen-bearing capri fig or wild fig (*see* HYMENOPTERA). The seedy fruits of the Smyrna-type fig are considered of finer flavor than the kinds without seed. In the United States, figs are grown primarily in California with some production in Arizona, Texas, and Florida. The annual crop value is approximately $8,500,000. *See* FRUIT (TREE). [C.A.S.]



Adriatic fig. (*From L. H. Bailey, ed., The Standard Cyclopedia of Horticulture, vol 2, Macmillan, 1937*)

Filariasis

The presence of any of the following hairlike nematodes in man: *Wuchereria bancrofti*, *W. pacifica*, *W. malayi*, *Onchocerca volvulus*, *Loa loa*, *Mansonella ozzardi*, *Acanthocheilonema perstans*, *A. streptocerca*. Adult worms of the first three inhabit the lymphatic channels; *O. volvulus* and *L. loa*, the subcutaneous connective tissue; and the last three, the body cavities.

Table 2. Important species

Scientific name	Common name or disease	Vector	Host
Wuchereria bancrofti	Elephantiasis	Mosquitoes	Man
W malayi	Human filariasis	Mosquitoes	Man
Loa loa	Eyeworm of man	Deerflies	Man
Onchocerca volvulus	Onchocerciasis	Blackflies	Man
O cervicalis			Horse
O gutterosa			Cattle
O gibsoni			Cattle
Dirofilaria immitis	Heart worm of dog	Fleas and mosquitoes	Dog
Stephanofilaria stilesi	Skin lesions of cattle	Unknown	Cattle
Elaeophora schneideri	Dermatiasis of sheep	Unknown	Sheep
Mansonella ozzardi	Mansonelliasis	Midge	Man
Acanthocheilonema perstans	Acanthocheilonemiasis	Midge	Man

filarial infection. Blood is collected, a stained smear is made and examined for microfilariae, or fluid is withdrawn from a nodule and examined in a similar manner. As the microfilariae are characteristic, a diagnosis may be made on their presence as well as supporting clinical symptoms of the

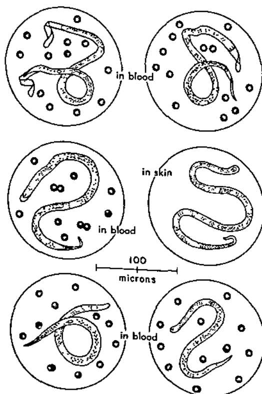


Fig. 3. Microfilariae of important human somatic nematodes. (*From Naval Medical School, Medical Protozoology and Parasitology, National Naval Medical Center, 1955*)

host. Figure 3 illustrates the important microfilariae of humans as they might appear in stained preparations. *See* DIPTERA; FILARIASIS; NEMATODA. [E.E.WE.]

Bibliography: D. L. Belding, *Textbook of Clinical Parasitology*, 2d ed., 1952; T. W. M. Cameron, *The Parasites of Domestic Animals*, 2d ed., 1952; T. T. Mackie, G. W. Hunter, III, and C. B. Worth, *A Manual of Tropical Medicine*, 2d ed., 1954

Filbert

Any plant belonging to the genus *Corylus* or the edible nut it produces. The terms hazel and hazelnut are also commonly used, but the officially recognized name is filbert. Eleven species of filberts [illegible] almost [illegible] 60 ft, and the Chinese species grows to a height of 120 ft. Filberts of Europe and the United States range from shrubs to small trees 25 ft in height. *See* NUT CROP CULTURE. [P.D.S.]

Filicales

These plants, commonly called true ferns, collectively make up the largest order of modern ferns with approximately 250 genera and 9000 species. Although well represented in the temperate regions, they reach their greatest development in the moist tropics. They vary in habit from small filmy structures to large treelike plants (Fig. 1). Many are epiphytic (live perched on other plants) and a number are climbing species. A few are aquatic. Perhaps the most striking species are the tropical tree ferns with their upright, unbranched stems and terminal clusters of large graceful leaves. The Filicales differ from the other fern orders, Ophioglossales and Marattiales, in being leptosporangiate, that is, the sporangium, or spore sac, arising from a single surface cell, and in having small sporangia with a definite number of spores. The wall of the sporangium is almost encircled with a ring of cells having unevenly thickened walls. This ring is called the annulus. When the sporangium is mature, the annulus, acting like a spring, causes the sporangium wall to rupture, thus discharging the spores. These plants are valued for their beauty and for the clues they give to the evolutionary history of the Filicales which extends back through the coal measures of the Paleozoic. *See* PALEOBOTANY.

Structure. The sporophyte is the conspicuous phase of the true ferns, and like other vascular plants, it has true roots, stems, and leaves (Fig. 2). In most ferns, especially those of the temperate regions, the mature stem is usually a creeping rhizome (underground stem) without aerial branches. However, in several species the stems may be branched, and in some they are erect. *See* STEM (BOTANY). Whereas in the tropics the leaves are usually persistent and evergreen, in temperate regions the leaves of most species die back each year

and are replaced by new ones the next growing season. *See* LEAF (BOTANY).

Characteristic of this order is the apparent uncoiling of the leaves from the base toward the apex, a phenomenon known as circinate vernation (Fig. 3). Few ferns have simple leaves, the common type being the compound pinnately divided leaves known as fronds. The leaflets of the fronds are called pinnae, and if they are further divided, the subdivisions are known as pinnules. The branching of the veins in the leaf is characteristically dichotomous and open



Fig. 1. *(Top)* Polypody fern on a cliff edge *(photograph by C. H. Jones). (Left)* The largest known specimen of the staghorn fern, an epiphyte on trees in the tropics. Photosynthesis occurs chiefly in the upright leaves. The rounded leaves pressed against the tree cover masses of roots. Reproductive structures develop on the pendant leaves *(photograph made on island near Brisbane, Australia, by C. J. Chamberlain, World Book Co.). (Top right)* A tree fern (U.S. Forest Service). *(Lower right)* The cultivated Boston fern and five of its many mutant varieties, all originated from mutations in vegetative cells. *(From E. N. Transeau, H. C. Sampson, and L. H. Tiffany, Textbook of Botany, rev. ed., Harper, 1953)*

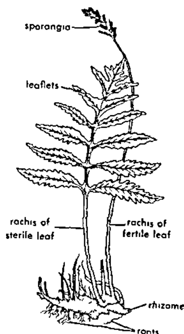


Fig 2. The sensitive fern (*Onoclea sensibilis*), a representative of the Filicinae. (*From W. W. Robbins, T. E. Weier, and C. R. Stocking, Botany: An Introduction to Plant Science, 2d ed., Wiley, 1957*)



Fig 3. Fern leaves pushing upward through a cement sidewalk. *Growth pressure may amount of hundreds of pounds to the square inch.* (*After G. E. Stone from E. N. Transeau, H. C. Sampson, and L. H. Tiffany, Textbook of Botany, rev. ed., Harper, 1953*)

The branching of the *adventitious* root is always ...

... display a large apical meristematic cell in both stem and root. *See* MERISTEM, APICAL; ROOT (BOTANY).

The internal structure of the blade of the leaf and of the root is very similiar to that of these organs in the seed plants. The *main* difference is the presence of large intercellular spaces in the fern leaf and the frequent lack of apparent distinction between the spongy and palisade cells of the mesophyll, possibly because most ferns grow in the shade. The vascular structure of the ferns *includes* four *types of stele* shown in Fig. 4: (1) The protostele, in which the phloem (food-conducting tissue) completely surrounds the xylem ...

(2) The amphiphloic siphonstele in which the xylem forms a cylinder enclosing the pith, with phloem bordering the xylem on both the inside and outside. (3) The ectophloic siphonostele having only external phloem surrounding the xylem which surrounds the pith. (4) The dictyostele which has the stele broken up into a number of vascular bundles either scattered or arranged in a circle (*see* STELE; VASCULAR BUNDLES). The dictyostele is the common type, with the amphiphloic siphonostele occurring rather frequently, and the other two types being comparatively rare. No cambium is present in the true ferns so that all the tissue is of primary origin (*see* PLANT TISSUE SYSTEMS). The development is mesarch (centripetal and centrifugal) since the metaxylem surrounds the protoxylem (*see* PLANT ANATOMY). The protoxylem consists of spirally thickened tracheids, the thickenings appearing as bands on the cell walls. Thick-walled sclerenchyma cells are numerous in most fern stems (*see* SCLERENCHYMA). Except in the few species which have a protostele, leaf gaps are present where the leaf (vascular) traces emerge from the stele.

Alternation of generations. The life cycle of the fern consists of two independent (self-sustaining) alternating generations (Fig. 5). The common leafy fern plant is the sporophytic (spore-producing) generation. When the mature spores are discharged and reach a suitable substrate, they germinate and produce a small, flat, green, heart-shaped structure known as the prothallium or gametophytic (gamete-producing) generation. The

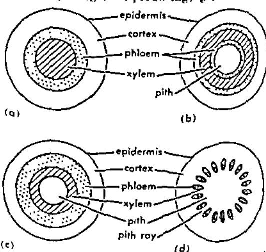


Fig. 4. Types of steles. (a) Protostele, with central rod of xylem surrounded by phloem, but with no pith. (b) Amphiphloic siphonostele with central pith and with a layer of phloem inside of the xylem as well as outside. (c) Ectophloic siphonostele with central pith, layer of xylem surrounded by phloem. (d) Dictyostele with central pith surrounded by xylem and phloem in the form of discrete vascular bundles. Bands of parenchyma cells known as pith rays are located between the bundles. (*From H. J. Fuller and O. Tippo, College Botany, Holt, 1949*)

gametophyte produces the sex organs antheridia (male) and archegonia (female). The gametes (sperm and egg) unite in fertilization and the resultant cell, or zygote, develops into the spore-bearing (sporophytic) fern plant.

In all ferns, the spores are produced in special multicellular organs known as sporangia. Except for a few genera, the sporangia are arranged in groups or clusters called sori (Fig. 6). These are

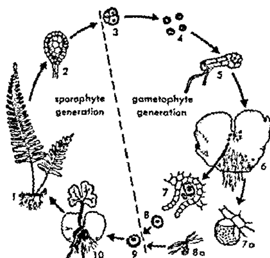


Fig. 5. The life history of a fern. 1, the fern plant, or sporophyte, bearing sori, or clusters of sporangia, on its leaves; 2, a sporangium; 3, a tetrad of young spores; 4, the four mature spores which have come from the tetrad shown in 3; 5, a spore germinating into a young gametophyte; 6, mature gametophyte, bearing sexual organs; 7, archegonium; 7a, antheridium, 8, egg cell, or female gamete; 8a, sperm, or male gamete; 9, fertilized egg; 10, young sporophyte growing out of a fertilized egg, the whole still attached to the remains of the gametophyte. (From E. W. Sinnott and K. S. Wilson, *Botany: Principles and Problems*, 5th ed., McGraw-Hill, 1955)

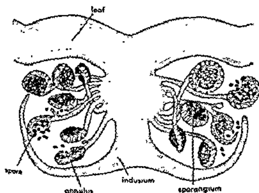


Fig. 6. Section through a fern leaf showing details of a sorus. (From W. W. Robbins, T. E. Weier, and C. R. Stocking, *Botany: An Introduction to Plant Science*, 2d ed., Wiley, 1957)

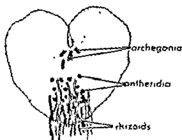


Fig. 7. Diagram of a fern prothallium. (After Gifford from W. W. Robbins, T. E. Weier, and C. R. Stocking, *Botany. An Introduction to Plant Science*, 2d ed., Wiley, 1957)

on the lower (abaxial) surface of the leaves or fertile fronds, either along the midrib of the pinnae, near the leaf margins, or scattered. Usually each sorus is covered by a flaplike structure called the indusium, which may be of various shapes and sizes. However, a few ferns have naked sori, and others have a false indusium formed by the folding or inrolling of the leaf margin. Varying degrees of specialization occur between the vegetative, or sterile, leaf and the reproductive, or fertile, leaf with most species having leaves able to perform both functions. As a result of meiotic divisions, each sporangium produces a number of tetrads of spores from the spore mother cells (*see* MEIOSIS). These spores are all alike (homosporous). When the spores are mature and conditions are favorable, the annulus ruptures the sporangium, throwing the spores some distance.

The spore is the first cell of the gametophytic generation. Each spore may germinate and produce a short filament of cells which eventually develops a triangular apical meristematic cell. By repeated cell division, the apical meristematic cell produces the heart shaped or lobed prothallium or gametophyte (Fig. 7). In most species the gametophyte is monoecious; that is, both male and female sex organs are on the same thallus. While still quite young, antheridia develop on the lower surface. The archegonia are produced later when the prothallium has matured and developed a thickened archegonial cushion just posterior to the growing point (meristem). Since gametophytes usually occur in various stages of development, it is likely that this difference in maturation time of the cells insures the crossing of gametes from different prothalli or gametophytes.

The antheridium is a rounded structure directly attached to the thallus or sometimes borne on a short stalk. Although each antheridium produces only a few sperms, the several antheridia present on each prothallus insures a large number of sperms. Each sperm is large, spirally coiled, and has many cilia.

The archegonia develop in such manner that when mature, the archegonial venter containing the egg is embedded in the prothallium whereas the neck cells project above the surface. The ma-

ture archegonium liberates a substance into the water which elicits a chemotactic response from the sperms, attracting them to the venter and the egg cell therein where fertilization is accomplished and a zygote is formed.

The ferns are characterized by regularity in the development of the embryo. A vertical division of the zygote followed by a transverse division produces a quadrat of cells. Each of these cells then divides to produce a group of eight cells in four pairs. The anterior pair of inner cells gives rise to the underground stem or rhizome, whereas the posterior pair produces the foot, a short-lived absorbing organ. The anterior pair of outer cells forms the first leaf and the other pair produces the primary root. The first leaves usually do not suggest the species, but these are followed by a series of transitional forms until finally the characteristic type of leaf is produced. The production of spores by the adult plant completes the sporophytic generation of the life cycle. Because ferns have long been of special interest to botanists, a great deal of morphological and paleobotanical evidence has been accumulated which strongly points to their origin from the Psilophytales (*see* PSILOPSIDA). There is also considerable evidence to show that some ancient group of the Filicineae may have given rise to some of the seed plants (Angiospermae) as exemplified by the Cycadofilicales and their modern descendants. *See* PTEROPSIDA.

[P.A V.]

Bibliography: H. C. Bold, *Morphology of Plants*, 1957; A. W Haupt, *Plant Morphology*, 1953; R. C. McLean and W. R. Ivimey-Cook, *Textbook of Theoretical Botany*, vol. 1, 1951; G. M. Smith, *Cryptogamic Botany*, vol. 2, 2d ed., 1955.

Filicineae

A class of the plant subphylum Pteropsida containing a large group of plants commonly called ferns. They are widely distributed throughout the world with their greatest development in the moist tropics. The stems, which may be unbranched or branched, frequently bear several large, spirally arranged leaves which are usually compound. *See* LEAF (BOTANY); STEM (BOTANY). In the majority of the species, the green leaves, sometimes called fronds, are the conspicuous organs of the sporophyte. The sporangia (spore sacs) are grouped in special structures called sori on the under margin or abaxial (lower) side of the leaves. Most of the sporangia are borne in groups. The fern sporophyte possesses true stems, leaves, and roots. *See* ROOT (BOTANY). Leaf and branch gaps both are present in the conducting cylinder or stele (*see* STELE).

Filicineae is an old class with a good representation in the Paleozoic flora (*see* PALEOBOTANY). Some of the plant fossils resemble contemporary living species. *See* MARATTIALES; OPHIOGLOSSALES; PTEROPSIDA.

See also COENOPTERIDALES. [P.A.V.]

Bibliography: *See* FILICALES

Filler

A material used to smooth wood surfaces before the application of finishing coats. Many woods, such as oak and mahogany, have large pores which interfere with the production of a smooth finish. Fillers are usually colorless materials that are formulated to harden rapidly with minimum shrinkage. The typical filler is made of silica or other hard pigment and dispersed in a fast-drying varnish. Filler is usually applied to the whole surface and, after partially setting up, is wiped to remove the excess from the high spots, leaving the pores filled level with the rest of the surface. Because the filler must be smoothed by sanding, it should harden enough to be sanded in a short time.

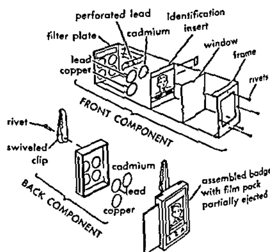
Fillers are usually colorless and are stained along with the rest of the wood. However, for special effects, colored fillers may be used which emphasize rather than minimize pores and grain.

Materials similar to putty but harder-drying which are used to fill cracks before surfaces are repainted, are sometimes referred to as fillers. The surfacers used to smooth dents in metals before finishing are also called fillers. *See* SURFACE COATING; WOOD FINISHING. [F.SD.]

Film badge

A device worn for the purpose of indicating the absorbed dose of ionizing radiation received by the person wearing it. A film badge usually is made of metal, plastic, or paper and is loaded with one or more film packets (ordinarily standard dental film packets, 1¾ by 1¼ in.). It is attached to the clothing by means of a clip or pin and is frequently combined with an identification badge. In some cases, it is made into a film ring or a film bracelet in order to indicate specifically the absorbed dose received by the hands or forearms.

The badge must contain one or more filters (usually made of cadmium, lead, silver, aluminum, or



Oak Ridge National Laboratory film badge.

copper) so that a comparison may be made of the relative blackening of the developed film from behind the various filters. This comparison reveals the extent of exposure to various types of radiation. A film badge usually contains a thin section or open window. The blackening of the developed film from this section gives a qualitative indication of the exposure to β-radiation. Some film badges contain an additional film having a thick emulsion that is used in estimating the neutron exposure. This is done by counting the density of proton recoil tracks in the developed film.

Film emulsions which permit a monitoring range from about 30 millirad to several thousand rad are available for use in film badges. For accurate results, control films and a set of calibrated films must be developed along with the monitoring films and the relative blackening measured with a densitometer. *See* DOSIMETER; MONITORING (IONIZING RADIATION). [K.Z.M.]

Filter, acoustic

A device employed to reject sound in a particular range of frequencies while passing sound in other ranges of frequencies. Acoustic filters are most often used in mufflers for automobile, diesel, and other reciprocating engines and on industrial machines such as air compressors. Acoustic filters have distinct advantages for these applications because they are usually constructed of steel or some other metal which gives long life with relatively little deterioration. No acoustical absorbing materials are generally employed in acoustic filters; hence they are generally considered reactive devices. Great ingenuity has been applied in combining reactive acoustical components to achieve the desired filter properties in a given space. For some examples of these complex configurations employed on reciprocating engines, *see* MUFFLER.

Filter design. The design of acoustic filters is usually carried out in direct analogy to the design of electrical filters by making use of the analogous relationships which exist between individual acoustical components and their electrical analogs. *See* DYNAMICAL ANALOGIES.

Direct analogy between electrical and acoustical elements in filter design usually assumes a lumped-circuit rather than a transmission line approach This assumption is valid for acoustical elements whose average linear dimensions are small compared to a wavelength of sound. This limits the validity of this approach, in general, to filters wherein the frequencies of interest usually lie below 1000 cps. When the dimensions of the acoustical components of a filter become of the order of ¼ wavelength or larger, electrical analogies can still be employed by resorting to the distributed parameter approach used in designing electrical filters at microwave frequencies. *See* FILTER, ELECTRIC.

Filter types. Generally, filters are classified as high pass, which reject low frequencies while passing high frequencies, or as low pass. Combinations of high- and low-pass filters forming a band-pass filter, which rejects frequencies above and below a specified frequency range, can also be constructed.

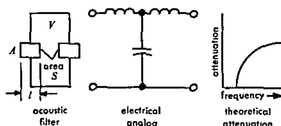


Fig. 1. Simple low-pass filter.

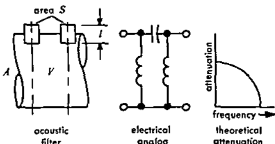


Fig. 2. Simple high-pass filter.

One of the simplest forms of a low-pass acoustic filter is shown in Fig. 1. This filter consists of two short tubes placed in the ends of a closed tank. Sound entering at A will pass through one tube and out the other only if the frequency is lower than the cutoff frequency f_c given by the equation

$$f_c = \frac{1}{\pi}\sqrt{\frac{c^2S}{l_eV}} \quad \text{cps}$$

where c = speed of sound in ft/sec; S = cross-sectional area of the tube in ft²; $l_e = l + 0.8\sqrt{S}$ = effective length of the tube in feet; and V = volume of the tank in ft³. The electrical circuit analogous to this filter and its theoretical attenuation characteristic are also shown in the figure.

A simple high-pass filter using the same basic elements as the low-pass filter is shown in Fig. 2, along with its analogous electrical circuit and theoretical attenuation characteristic. Sound entering the open-ended tank at A will be transmitted through the filter only if the frequency is higher than the cutoff frequency given in the preceding equation.

Attenuation characteristics. The theoretical attenuation characteristics shown in Figs. 1 and 2 assume that the filter is connected at each end into its image impedance z_i. For the filters indicated, this impedance is given by the equation

$$z_i = \frac{p}{Sv}$$

where p = sound pressure, S = cross-sectional area, and v = sound-particle velocity. In general,

it is not possible to provide such an impedance match, and therefore there is a consequent loss in performance for the filter. Acoustical dissipation, which is neglected in the theoretical analysis, will also tend to degrade the filter performance. *See* IMPEDANCE, ACOUSTIC. [W.J.G.]

Bibliography: W. P. Mason, *Electromechanical Transducers and Wave Filters*, 2d ed., 1948.

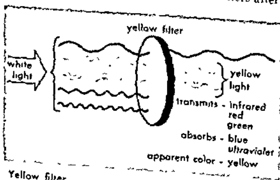
Filter, color

An optical element that partially absorbs incident light, often called a light filter. The absorption may be selective with respect to wavelength, in which case the color filter purifies the incident light in the sense of making it more nearly like that of a part of the pure spectrum On the other hand, the absorption may be nonselective, in which case the color filter merely reduces the light flux without change in spectral composition. Color filters are used in photography, in optical instruments, and in illuminating devices to control the amount and spectral composition of light.

Color filters are made of glass for maximum permanence, of liquid solutions confined to cells with transparent faces for maximum flexibility of control, and of dyed gelatine or plastic for maximum convenience combined with excellent flexibility and often satisfactory permanence. Plastic color filters are supplied cemented between glass sheets for greater toughness than glass and greater scratch resistance than bare plastic.

Color filters are sometimes classified according to the type of their spectral absorption, for example, neutral (spectrally nonselective), diffuse-cutting (gradual change in absorption with wavelength), sharp-cutting (absorbing one end of the spectrum and transmitting nearly all of the remainder), band pass (absorbing all incident energy except that of a stated wavelength range), and narrow-band pass (absorbing all incident energy except that of a narrow wavelength range). A sharp cutting yellow filter is shown in the illustration.

... band of wavelengths nearly the same refractive index as the particles themselves. These filters, called Christiansen filters after their discoverer C. Christiansen, do not absorb the energy of unwanted wavelength, but refract it out of the direct beam.



Yellow filter.

Narrow-band-pass filters may be made by depositing on glass two thin metallic films separated by a thin dielectric layer. These filters, called interference filters, prevent the passage of energy of unwanted wavelength by reflecting it. *See* INTERFERENCE FILTER, OPTICAL; *see also* ABSORPTION (ELECTROMAGNETIC RADIATION); COLOR. [D.B.J.]

Bibliography: H. S. Denmark and W. M. Cady, Optimum grain size in the Christiansen filter, *J. Opt. Soc. Am.*, 25:330, 1935; H. P. Gage, Glass color filters for special applications, *J. Opt. Soc Am.*, 27:159, 1937; L. N. Hadley and D. M. Dennison, Reflection and transmission interference filters, *J. Opt. Soc. Am.*, 37:451, 1947, 38:483, 546, 1948; D L. MacAdam, Colorimetric specifications of Wratten light filters, *J. Opt. Soc. Am.*, 35:670, 1945; J. Strong, *Concepts of Classical Optics*, 1958.

Filter, electric

A transmission network used in electrical systems for the selective enhancement of a given class of input signals. While the term is sometimes used in a broad sense to include all networks (nonlinear as well as linear) which accomplish any type of enhancement, the usual connotation is a network of essentially linear elements. In this case the enhancement is a result of the selective attenuation of undesired frequency components of the input wave relative to the frequency components of the desired signal. The effect of such a filter is most easily described in terms of the associated transmission function which is the ratio of the magnitude of the output to the input signals for steady-state operation. This function indicates which bands of frequency components are transmitted (the pass bands), and which are attenuated (the stop bands). In terms of the transmission function, the effect is quite analogous to that of an optical filter which transmits incident light in certain portions of the frequency spectrum with little attenuation, while light of other frequencies is attenuated almost completely.

Classification of filters. Filters are classified into a number of nonexclusive categories according to the elements they contain, the type of transmission function, the design criteria employed, and the intended use. The first classification leads to a basic designation as active or passive; this depends on whether the linear elements comprising the filter are all passive resistors, inductors, and capacitors or whether active elements, such as vacuum tubes and transistors, are included.

Another important classification is based on the pass and stop bands of the transmission function. Most filters fall into one of four categories.

Low-pass filter. This is a filter which transmits more or less uniformly in a frequency range from zero frequency (dc) to some finite cutoff frequency. Frequency components higher than the cutoff frequency are largely attenuated.

High-pass filter. This is a filter which transmits more or less uniformly in a frequency range from some finite cutoff frequency to the highest frequencies of interest. Frequency components lower than the cutoff frequency are attenuated.

Band-pass filter. This is a filter which transmits more or less uniformly in a frequency range between lower and upper cutoff frequencies. Frequency components outside the pass band are attenuated.

Band-stop filter. Also known as a band-elimination or band-rejection filter, this is a filter which transmits more or less uniformly at all frequencies of interest except for a stop band within which frequency components are largely attenuated.

Transmission characteristics. Figure 1 shows idealized transmission functions corresponding to the above classifications of filters. The uniform or flat transmission in pass bands, zero transmission in stop bands, and the implied discontinuous transitions between pass and stop bands are not attainable with physically realizable elements, although approximate analyses are often based on such ideal filters. Figure 2 depicts practical specifications for the transmission function of a low-pass filter. The tolerances in the pass and stop bands are indicated by T_a and T_a respectively. Furthermore, the variation from pass to stop band is not discontinuous but occurs over a finite transition region ($\omega_a - \omega_b$). In the transition region, the slope or rate of cutoff is usually specified in db per octave since the logarithm of the transmission function falls at approximately a constant slope when plotted as a function of the logarithm of the frequency. The band edge ω_b is the point where the transmission function falls below the pass-band tolerance T_a. The cutoff frequency usually refers to the frequency $\omega_c = \sqrt{\omega_a\omega_b}$ although this definition is not unique.

Design specifications are not always limited to the quantities above. For example, characteristics



Fig. 1. Transmission functions of ideal filters.

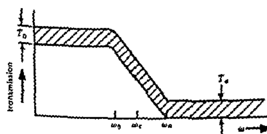


Fig. 2. Tolerances in the specification of a transmission function.



Fig. 3 Commonly used pass-band shapes.

in the transmission region might be prescribed in greater detail as in Fig. 3 where the maximally flat, Butterworth, or equal ripple, Chebycheff, shapes are illustrated. Filters are also classified in accordance with the design procedure used, each having a characteristic pass-band shape.

Another important category includes those filters which are not based on the concept of pass and stop bands. Instead the effect of the filter is described in terms of a system parameter, such as output signal-to-noise ratio.

Filter applications. Filters are used for a variety of purposes in nearly every type of electronic communication and control equipment. In conventional radio communication systems the frequencies present in the signal generated at the transmitter are confined to a relatively narrow band near the carrier frequency. A band-pass filter is incorporated in the receiver so that these frequency components are selectively amplified while undesired signals and noise in surrounding frequency ranges are attenuated to minimize their effect at the receiver output. Similarly, the audio and video signals which comprise a complete television signal are separated for further signal processing by a system of frequency selective filters in a television receiver. The bandwidths of filters used in communication systems vary from a fraction of a cycle per second to many megacycles per second depending on the application.

Sometimes, as in the case of interference from 60-cps power lines in an audio system, the exact frequency of a spurious noise signal is known. Usually such interference is from an odd harmonic of 60 cps, for example 300 cps. A sharply tuned band-rejection filter, designed to attenuate the 300-cps noise signal, is used to minimize the effect at the output. A high-pass filter with cutoff frequency greater than 300 cps would also remove the interference, but some of the lower frequency components of the audio signal would be lost as well.

Filters are employed to isolate various sections of a complete system and thus to prevent undesired interactions. For example, the insertion of low-pass decoupling filters between each of several amplifier stages and a common power supply reduces interaction due to the common power supply impedance. Perhaps, the most common application of filters in electronic circuits is the use of power supply filters to remove the ac components pre... the output of a rectifier. Here the desired ...

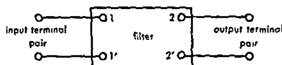


Fig. 4. A typical two-terminal-pair network.

is simply the dc component, and the ac components are largely attenuated by a low-pass filter

In each of the above examples, the action of the filter is easily understood in terms of the ideal filter concept of transmitting energy in a given band of frequencies while attenuating other frequency components. A more sophisticated type of filtering can be achieved when more information about the desired signal and the undesired interference is available. Even in the case where the spectra of the signal and interference occupy the same frequency band it is possible to design filters which selectively enhance the desired signal provided that the spectral shapes of interference and signal are known. In general the best filters in such an application are not ideal filters but filters with transmission functions which vary continuously *with frequency*.

Filter design. A filter is usually a two-terminal-*pair network* with one pair designated as input terminals and the other pair as the output, as depicted in Fig. 4. Many passive filters are constructed of reactive (L and C) elements and terminated in a resistor. Sometimes resistors are incorporated in the body of the network. LC filters transmit with minimum loss in the pass band but have the disadvantage of requiring inductive elements which may be expensive. When amplification is available it is often more economical to use only R and C elements.

A decoupling filter (low-pass) is an example of an RC *filter*. *Referring to* Fig. 5, a dc voltage E_1 plus a superimposed ac voltage e_1 are assumed to be applied to the input of this two-terminal-pair filter. The output voltages are readily obtained by an application of basic circuit laws. The output dc voltage E_2 is

$$E_2 = E_1 - IR$$

where I is the dc current through the filter and R is the resistance element of the filter. *Thus for* direct current there is a loss in voltage which may be kept small by proper choice of R. The magnitude of the ac voltage e_2 passed by the filter is given by

$$|e_2| = |e_1| \frac{1}{\sqrt{1 + (\omega RC)^2}}$$

provided that the equivalent load resistance of the amplifier is considerably greater than $|1/\omega C|$, as is usually the case. When $\omega RC = 1$ then

$$|e_2| = |e_1| \frac{1}{\sqrt{2}} = 0.707|e_1|$$

The frequency $f_1 = 1/2\pi RC$ for which this occurs is said to be the half-power frequency. This is the frequency for which the output $|e_2|$ is down -3db of the input $|e_1|$. Obviously for any frequency greater than f_1 the attenuation is correspondingly more than 3db. *See* DECIBEL.

A second example of a common low-pass filter is that used for smoothing the dc output voltage of a rectifier (see Fig. 6). As with the decoupling network, the dc voltage applied to the filter network passes through it attenuated only by the Ir drop across the inductance L (which has a small resistance r). This resistance may be made quite low; therefore the LC filter possesses one important advantage over the RC filter in that the dc voltage attenuation is very small.



Fig. 6 An L-section LC filter (low-pass) for power supplies

For an ac voltage e_1 applied to the filter, the magnitude of the output ac voltage is

$$|e_2| = |e_1| \frac{1}{1 - \omega^2 LC}$$

provided that the load impedance is large com- [illegible] operated [illegible] voltage for the lowest frequency involved (120 cps for a full-wave 60-cps rectifier). This requires selection of a large filter capacitor C to go with available inductors L. *See* POWER SUPPLY, ELECTRONIC; RECTIFIER.

Many network configurations are used in the design and construction of filters. An unbalanced *network* has one grounded terminal common to input and output and is usually preferred for practical reasons. The recurrent ladder structures shown in Fig. 7 are frequently used for passive filters. Qualitatively, the filtering effect is achieved by choosing Z_a and Z_b such that $|Z_a| \gg |Z_b|$ in the frequency range which is to be a stop band and $|Z_a| \ll |Z_b|$ in the desired pass band.

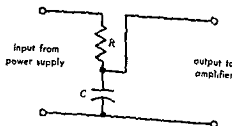


Fig. 5. A decoupling network between power supply and amplifier stage.

Network elements

	low pass	band pass	band stop	high pass
Z_a	Ls	$Ls+\frac{1}{Cs}$	$\frac{1}{Cs+\frac{1}{Ls}}$	$\frac{1}{Cs}$
Z_b	$\frac{1}{Cs}$	$\frac{1}{Cs+\frac{1}{Ls}}$	$Ls+\frac{1}{Cs}$	Ls

Note: $s = j\omega$ for sinusoidal steady-state operation

The table indicates the simplest choice of network elements which have the required frequency dependence for the various types of pass bands. Analysis and design procedures for recurrent networks are considerably simplified if it is assumed that the successive sections are identical and the length of the ladder is infinite. This simplification accounts, in part, for the frequent use of constant-k filters which are discussed below. Recurrent ladder networks in which the successive sections are not identical are used to obtain filters with desirable transmission functions of the maximally flat and equal-ripple type. In both cases the approximation to an ideal filter is improved as the number of sections is increased.

Network configurations other than the ladder structure permit greater flexibility in the characteristics of the transmission functions. In particular, zero transmission at a specified frequency may be achieved by structures having two paths between input and output. The values of the network elements are chosen in such a way that the outputs due

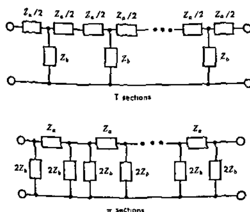


Fig. 7. Two forms of recurrent ladder networks.

to each of the paths precisely cancel at the specified frequency. The twin-T filter, Fig 8, has this property and is frequently used to eliminate a specified frequency component.

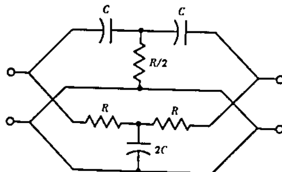
The above discussion suffices only to describe the mechanism of filtering. Advanced mathematical techniques are required to approximate a specified filter response and to determine the network structure and numerical values of the elements. The solution to the design problem is not unique in general, and practical considerations usually dictate which of several possible designs is most desirable.

Types of design. The following sections discuss eleven types of filter design.

Constant-k filter. The constant-k type is a recurrent filter network, Fig. 7, in which each section has the property

$$Z_aZ_b = k = R_o^2$$

where Z_a and Z_b are purely reactive (no resistive elements) and k is a constant with the dimensions of (ohms)^2. Constant-k filters are often used to approximate ideal filters with uniform pass bands The relatively simple design equations for a single section require only the specification of the cutoff frequencies and the terminating resistance R_o in the pass band. The formulation of these design equations is based on an infinite recurrent structure in which successive sections are identical. In practice, however, the filter consists of only a finite number of sections terminated in a resistor which modifies the theoretical transmission function slightly. Restricting the filter network to a recurrent structure simplifies the design procedure, but does not necessarily lead to the best performance for a network of a given number of elements. Since Z_a and Z_b are required to be pure reactances, the effect of the finite loss associated with actual



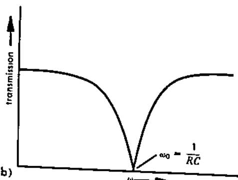


Fig. 8 Twin-T rejection filter. (a) Network. (b) Transmission function.

physical elements can not easily be taken into consideration.

The equation above requires that Z_a and Z_b have inverse variation with frequency as indicated in the table. The pass and stop bands of constant-k filters are determined from the following conditions:

$$-4 \leq Z_a/Z_b \leq 0 \quad \text{(pass band)}$$
$$Z_a/Z_b \leq -4 \quad \text{(stop band)}$$
$$Z_a/Z_b = -4 \quad \text{(impedance ratio at cutoff frequency)}$$

The terminal impedance of the filter is relatively constant in the pass band and approximately equal to R_o. Thus for a constant-k low-pass filter with a termination R_o and cutoff frequency ω_c, the following equations determine the network parameters:

$$Z_aZ_b = L/C = R_o^2 \qquad -\omega_c^2LC = -4$$

The transmission function of a constant-k filter is dependent on the number of recurrent sections used. In general, increasing the number of sections results in a more uniform response in the pass band, more attenuation in the stop band, and a narrower transition region between pass and stop bands. As a rough approximation, each additional section increases the rate of cutoff by 12 db per octave. Differences in the transmission properties for the two recurrent structures in Fig. 4 are minor.

M-derived filter. This filter is derived from a constant-k structure by modifying one or more of the sections. A modification of the first and/or last section is often made to improve the impedance match between the filter and the terminating resistance. In addition, interior sections of a constant-k filter are sometimes replaced by m-derived sections to achieve a sharper cutoff, that is, a narrower transition region between pass and stop bands. In modifying a constant-k, T section, Z_a and Z_b are replaced by

$$Z'_a = Z_a \qquad Z'_b = \frac{Z_b}{m} + \frac{(1-m^2)Z_a}{4m} \qquad 0 < m < 1$$

where m is a design parameter. The choice $m = 0.6$ is used when the m-derived section is introduced to improve the impedance match between the filter and its termination.

The impedance Z'_b becomes zero at a frequency outside the pass band but near the associated cutoff frequency. The exact frequency of this null can be adjusted by choice of the design parameter m, and this property is often used in modifying constant-k sections to achieve a greater rate of cutoff, that is, a narrower transition region as in Fig 9. While this improvement is gained at the expense of less attenuation at frequencies remote from the cutoff, sufficient attenuation in this region is usually provided by the remaining unmodified constant-k sections in the composite filter. The complete filter including constant-k and m-derived sections has a nonrecurrent structure.

Active filter. An active filter incorporates active as well as passive elements. Active elements may be introduced to achieve a reduction in the size or expense of the filter or to provide gain in the pass band. Feedback is frequently employed to obtain desired results.

Image parameter filter. This filter is designed by a method using image impedance and image transfer functions as the fundamental network functions. Constant-k and m-derived filters are special cases.

Matched filter. A filter that is matched to the characteristics of an input signal of known time dependence is called a matched filter. The term is also applied to describe an impedance match between a filter and its termination. A matched filter has the following property: when the input consists of noise in addition to the desired signal, the

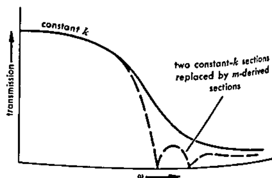


Fig 9. Effect of modifying a constant-k filter with two m-derived sections.

output signal-to-noise ratio is the maximum that can be obtained by any linear filter. Accordingly, such filters are used (in radar and similar systems) to optimize the detection of weak signals. The impulse response of a matched filter is simply related to the signal time function when the spectrum of the interference is uniform.

Mechanical filter. This type of filter is a mechanical transmission device in which mass or moment of inertia and elastic compliance are the frequency-sensitive elements. Mechanical filters may be coupled to electric networks by means of electromechanical transducers. Mechanical filters are sometimes used to achieve a reduction in equipment size, particularly in low-frequency applications or in applications where a large number of low-loss coupled resonant circuits are required. The low loss of some resonant mechanical systems (especially quartz crystals) is also an important consideration in their use as filters in frequency standards.

Microwave filter. The microwave filter is one designed to operate at frequencies so high that the physical dimensions of the filter are an appreciable fraction of a wavelength. Such filters are constructed as integral parts of two-conductor transmission lines or wave guides. Inductors and capacitors are replaced by short-circuited sections of transmission line or wave guide and, in wave guides, by resonant structures and by various modifications of the uniform guide by the insertion of posts, irises, and similar discontinuities. *See* MICROWAVE TRANSMISSION LINES; WAVE GUIDE.

Optimum filter. In common usage an optimum filter is one in which the mean square value of the error between a desired output and the actual output is at a minimum. While other criteria for defining optimum are possible, the above condition leads to meaningful results and, more significantly, leads to a class of soluble design equations. This approach to optimum filter design was originally developed by N. Wiener. Optimum filters are used in predicting the future values of the input, or any linear combination of the input and its derivatives, and in minimizing the mean square error between the output and input signal in the presence of interference. In all cases the power spectra of the desired signals and interference must be postulated

Smoothing (averaging, integrating) filter. This term is used for any low-pass filter Since the filter attenuates the fluctuations (ac components) present at the input, the output is smoothed and tends to respond to the average value of the input. Such filters are frequently used to reduce the effect of noise in physical measurements and are common in electronic power supplies.

Stagger-tuned filter. A filter consisting of a cascade of amplifier stages with tuned coupling networks is called a stagger-tuned filter. The resonant frequency and bandwidth of the individual resonant circuits may be easily adjusted to achieve an over-all transmission function of desirable (maximally flat or equal ripple).

Tapped delay-line filter. A tapped delay-line filter is one in which the output is a linear combination of signals obtained from taps on a delay line or transmission line In the latter case filter characteristics may be obtained which are not easily approximated by *LCR* networks. Tapped delay-line filters are particularly useful as matched filters for pulse-type signals such as those used in radar, which are essentially discontinuous time functions. [R.A.J.; W.R.L.]

Bibliography: N. Balabanian, *Network Synthesis*, 1958; H. W. Bode, *Network Analysis and Feedback Amplifier Design*, 1945; W. Cauer, *Synthesis of Linear Communication Networks*, vols 1–2, 2d ed., 1958; E. A Guillemin, *Communication Networks*, vol. 2, 1933; E. A. Guillemin, *Synthesis of Passive Networks*, 1957; W R. Lepage and S. Seely, *General Network Analysis*, 1952; M. B. Reed, *Electric Network Synthesis, Image Parameter Method*, 1956.

Filtration

The removal of solid particles from a fluid by forcing the fluid through a filter medium, or septum, on which the solids deposit. Industrial filtrations range from simple straining to difficult separations conducted in complicated equipment. The fluid may be liquid or gas; the solid particles may be coarse or fine, rigid or plastic, crystalline or colloidal The ratio of solid to liquid may be large or small. Sometimes the liquid is valuable, sometimes the solid, and sometimes both. Filtration may be done hot or cold, under pressure, or in a vacuum. Often it is necessary to wash the solid remaining on the septum to recover or eliminate soluble materials retained in the pores of the cake when filtration is complete. Because of these variations in the substances filtered and in the conditions of filtration, many different kinds of filters are in use.

Gas-solid filters. These are used to remove solid particles from a gas, as in the cleaning of dusty air. The dust-laden gas is passed through bags made of cloth or of special filter paper, or through layers of specially prepared fibers in which particles of solid are trapped. The filter of a cigarette is a little porous plug filter, and the dust collector of a household vacuum cleaner is a small bag filter. In industrial filters, automatic methods are provided for periodic purge of the retained solids. Bag filters remove particles ranging in size from 0.1–100 μ (approximately 25 μ equal 0.0001 in.).

Solid-liquid filters. These are used in the processing of chemicals, metallurgical sludges, and food products, and in water purification. In industry the liquid is ordinarily water or aqueous solution, but it may be any solvent. A pressure difference must be established across the filter to force the liquid through the septum and the cake.

Free-settling sludges may also be removed from liquids by simple settling. *See* SEDIMENTATION (INDUSTRIAL). Centrifugal force may be used [illegible]ering crystals from a viscous mother liquor [illegible]RIFUGATION).

Two main types of solid-liquid filters are used. Clarifiers are used to remove small volumes of solid from large volumes of liquid. The solid is usually an impurity to be discarded and may amount to only a few ounces in a thousand gallons of liquid (*see* CLARIFICATION). Cake filters are used when the solid-liquid ratio is large. Slurries containing as much as one-third or more of their volumes as solids may be so filtered. The usual solid content is in the range of 0.5–10% by volume.

Large volumes of water may be filtered through beds of sand In gravity filters, which are used in water purification, the weight of water on the filter bed provides sufficient pressure to operate the filter. Sometimes the sand bed is placed inside a pressure vessel and the liquid forced through by an external pump The sand bed may be backwashed periodically to cleanse it of accumulated solids.

Cake filters. These are used to remove from their slurries, solids such as dyes, fine crystals, metallurgical slimes, sludges obtained in the purification of sugar liquors, and other precipitates. They operate either by applying pressure to the slurry on the upstream side and discharging filtrate at atmospheric pressure, or by feeding the filter at atmospheric pressure and applying a vacuum on the downstream side. In the second method, the maximum obtainable pressure drop over the filter is 1 atm. In some applications, this is insufficient to give a practicable filtration rate, and pressure filters must be used. Most cake filters include means for washing the cake.

Filter presses. These constitute a versatile class of pressure filters One form, the plate-and-frame press, is shown in Fig. 1. This filter consists of a number of rectangular vertical plates, alternating with empty frames, assembled on horizontal rails. Heads, one fixed and one movable, hold the plates and frames together. Both sides of each plate and the inner surfaces of the [illegible]

tion area. [illegible]

over each [illegible]

plates and frames are compressed by a powerful screw, so that the cloths also act as gaskets to prevent filtrate leaking between the plates.

Channels lead into each frame from the slurry inlet at the fixed head. Channels for filtrate are cored into each plate, pass through a corner of

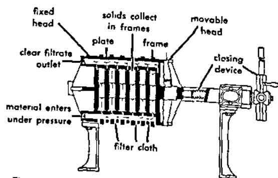


Fig 1. Corner feed, closed-discharge filter press. (From T. Shriver and Co.)

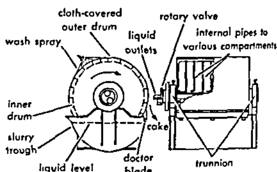


Fig 2 Continuous vacuum rotary filter. (*From W. L. McCabe and J. C. Smith, Unit Operations of Chemical Engineering, McGraw-Hill, 1956*)

each frame, and direct the filtrate into the outlet which is also in the fixed head. The pressure on the slurry forces the filtrate through the cloths. Filtrate collects on the plate surfaces and leaves the press through the filtrate outlet; cake collects in the frames. When the frames are full of solid, the filtration is stopped, the plates released by backing away the screw and sliding them apart, and the cake removed. The plates are then closed for another cycle.

Many kinds of plate-and-frame presses are available. They may be equipped for washing with special washing plates, which allow passage of wash water through the cake. Wash passes entirely through the cake from one plate to the next. The wash is collected on each alternate plate and guided to the fixed head through a separate channel

Rotary vacuum filters. These are used on free-filtering solids and operate continuously. The type shown in Fig. 2 is a cloth-covered horizontal drum, partially immersed in the slurry. The periphery of the drum consists of a number of individual filter elements, each connected to one of the trunnions (drum supports) by pipes. The trunnion contains a special valve, which times and controls the action of each filter element. The valve acts to apply suction to each element from the time it enters the slurry until just before it reaches the doctor blade It then directs compressed air to the element until the element again submerges. Suction is then reestablished and the element reenters the slurry.

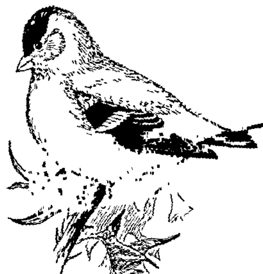
Each filter element draws filtrate during immersion and accumulates cake. The filtrate flows to the trunnion under the action of the vacuum and out of the filter. As the element emerges from the slurry, the filtrate remaining in the cake is removed by suction. Then wash water is sprayed on the cake and drawn out through the trunnion. The valve directs the wash into a channel separate from the filtrate draw-off. The cake is then sucked dry of wash and doctored off the filter, with assistance from compressed air. The filter element reenters the slurry, and the cycle is repeated. *See also* SEPARATION (MECHANICAL). [W.L.M.]

Bibliography: G. D. Dickey and C. L. Bryden, *Theory and Practice of Filtration*, 1946; W. L.

McCabe and J. C. Smith, *Unit Operations of Chemical Engineering*, 1956; J. H. Perry (ed.), *Chemical Engineers' Handbook*, 3d ed., 1950.

Finch

Any of numerous sparrowlike birds. There are finches throughout the world, as the name is applied to various groups of stout-billed, seed-eating birds. Some, such as the Java finch, are popular in collections of caged birds. The American goldfinch is a brilliant bird; others are drab in color.



The goldfinch, *Spinus tristis*; length to 6 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Besides the goldfinches, the name finch in America is usually applied to one of the species of the genus *Carpodacus*, family Fringillidae, a Holarctic group of 6 species, 3 of which are found in the United States. All are similar, brown, streaked sparrows, about as large as the English sparrow. The males are richly suffused with reddish feathers which are brightest on the rump, breast, and head.

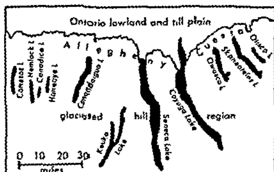
The house finch, *C. purpureus*, nests across North America in the coniferous forests. Its flocks are prominent in the spring migration. *C. mexicanus* is a western species which ranges from Oregon to southern Mexico and eastward to the Great Plains. *C. cassinii*, Cassin's purple finch, occurs in the coniferous forests of the western mountains from British Columbia to Baja California. See PASSERIFORMES; SPARROW. [J.O.B.]

Fine structure (spectral lines)

A term referring to the closely spaced groups of lines observed in the spectra of the lightest elements, notably hydrogen and helium. The components of any one such group are characterized by identical values of the principal quantum number n, but different values of the azimuthal quantum number l and the angular momentum quantum number j (see QUANTUM NUMBERS). According to P.A.M. Dirac's relativistic quantum mechanics, those energy levels of a one-electron atom that have the same n and j coincide exactly, but are displaced from the values predicted by the simple Bohr theory by an amount proportional to the square of the fine structure constant α (see ATOMIC STRUCTURE AND SPECTRA; QUANTUM THEORY, RELATIVISTIC). The constant α is dimensionless, and nearly equal to $1/137$. Its value is actually $0.00729729 \pm 0.00000003$. In 1947 deviations from Dirac's theory were found, indicating that the level having $l = 0$ does not coincide with that having $l = 1$, but is shifted appreciably upward. This is the celebrated Lamb shift, named for its discoverer, Willis Lamb, Jr. Modern quantum electrodynamics accounts for this shift as being due to the interaction of the electron with the zero-point fluctuations of the electromagnetic field. See QUANTUM ELECTRODYNAMICS. [F.A.J.]

Finger lakes

Long, narrow, comparatively straight lakes best developed in the north border of the Appalachian Plateau in western New York. These lakes strikingly resemble fiords; the bottoms of some are below



Sketch map of the Finger Lakes in the glaciated hill region of western New York. (From V. C. Finch et al., Elements of Geography, 4th ed., McGraw-Hill, 1957)

sea level. This fact and the sharp contrast with the smooth slopes of the uplands prove that preglacial or interglacial valleys were deepened by glacial erosion. Some tributary valleys hang. Localization of ice erosion was caused by concentration of flow and pressure melting bringing about plucking under thick ice. See FIORD; GLACIATED TERRANE. [F.T.T.]

Fingerprint

Distinctive ridges that appear on the bulbs on the inside of the end joints of the fingers and thumbs. These ridges have definite contours and appear in three general pattern types, the arch, the loop, and the whorl, each with general and specific variations of the pattern, dependent upon the shape and relationship of the ridges.

Fingerprints are an infallible means of identification. In addition to their value in the apprehension of criminals, fingerprints can ensure personal identification for humanitarian reasons, such as in cases of amnesia, missing persons, or unknown

deceased. Fingerprints are invaluable in effecting identifications in tragedies such as fire, flood, and vehicle crashes. In criminal matters, besides establishing the identity of the arrested person, fingerprint records provide a history of known offenders, or indicate when a person is a first offender. The vast majority of fingerprints maintained in the Identification Division of the Federal Bureau of Investigation of the United States, the largest repository of fingerprints in the world, are for civil records.

Patterns. Fingerprints fall into three general types of patterns. Each group bears the same general characteristics or family resemblance. The three general pattern types may be further divided into subgroups by means of smaller differences existing between the patterns in the same general group. The arch group includes the plain arch and the tented arch. The loop group includes the radial and ulnar loops. The whorl group includes four types of whorls, the plain whorl, central pocket loop, double loop, and accidental whorl.

Pattern area. The pattern area is that part of a fingerprint in which appear the cores, deltas, and ridges that are used for classification. The pattern areas of loops and whorls are enclosed by type lines, which may be defined as the two innermost ridges which start parallel, diverge, and surround or tend to surround the pattern area.

Focal points. Within the pattern areas of loops and whorls are the focal points which are used in the detailed classification. These points are called the delta and the core. The delta is that point on a ridge at or in front of and nearest the center of the divergence of the type lines. A core may be defined as that point on a ridge which is located in the approximate center of the finger impression.

Loop. The loop (Fig. 1) is that type of pattern in which one or more of the ridges enter on either side of the impression, recurve, touch or pass an imaginary line drawn from the delta to the core, and terminate or tend to terminate on or toward the same side of the impression from which such ridge or ridges entered. In the loop-type pattern the ridges intervening between the delta and core are counted.

The terms "radial" and "ulnar" are derived from the radius and ulnar bones of the forearm. The classification of loops is determined by the way the loops flow on the hand. Loops which flow in the direction of the ulnar bone, toward the little finger, are called ulnar loops and those which flow in the direction of the radius bone are called radial loops.

Arches. In plain arches (Fig. 2*a*) the ridges enter on one side of the impression and flow or tend to flow out the other side with a rise or wave in the center.

A tented arch (Fig. 2*b*) is that type of pattern which possesses either an angle, an upthrust, or two of the three basic characteristics of a loop. That is, the tented arch may lack a delta, ridge count or recurve, and type lines which are characteristics of the loop.

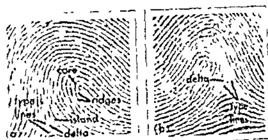


Fig. 1. (*a*) Twelve-count ulnar loop. (*b*) Radial loop (*Federal Bureau of Investigation*)

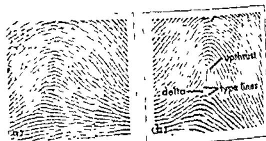


Fig. 2 (*a*) Plain arch. (*b*) Tented arch. (*Federal Bureau of Investigation*)

Whorl. The whorl is that type of pattern in which at least two deltas are present with a recurve in front of each. This definition is very general. The pattern may be subdivided into four subgroups for extension purposes in large groups where whorls are predominant.

Plain whorl. This pattern is the simplest and most common of the whorl subdivisions. It has two deltas and at least one ridge making a complete circuit, which may be spiral, oval, circular, or any variant of a circle. An imaginary line drawn between the two deltas must touch or cross at least one of the recurving ridges within the inner pattern area.

Central pocket loop. This pattern is that type of whorl which has two deltas and at least one ridge making a complete circuit. An imaginary line drawn between the two deltas must not touch or cross any of the recurving ridges within the inner pattern area. In lieu of a recurve in front of the delta in the inner pattern area, an obstruction at right angles to the line of flow will suffice. The inner line of flow is determined by drawing an imaginary line between the inner delta and the center of the innermost recurve or looping ridge.

Double loop. This pattern consists of two separate loop formations and two deltas.

Accidental whorl. This is a pattern consisting of a combination of two "different" types of pattern, with the exception of the plain arch, with two or more deltas; or a pattern which possesses some of the requirements for two or more different types; or a pattern which conforms to none of the definitions (Fig. 3).

The technique of whorl tracing depends upon the establishment of the focal points, the deltas. Every whorl has two or more deltas. When the deltas have

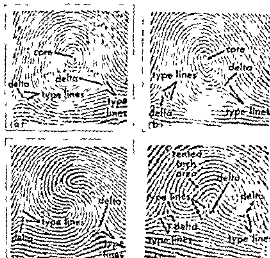


Fig. 3. (a) Plain whorl, O tracing. (b) Central pocket loop, O tracing. (c) Double loop, I tracing (d) Accidental whorl, O tracing. This is a loop over a tented arch. (Federal Bureau of Investigation)

been located, the ridge emanating from the lower side or point of the extreme left delta is traced until the point nearest or opposite the extreme right delta is reached. The number of ridges intervening between the tracing ridge and the right delta is then counted. If the ridge traced passes inside of (above) the right delta, and three or more ridges intervene between the tracing ridge and the delta, the tracing is designated as "inner" or I. If the ridge traced passes outside (below) the right delta, and three or more ridges intervene between the tracing ridge and the right delta, the tracing is designated as "outer" or O. All other tracings are designated as "meeting" or M. When the ridge traced ends abruptly, the tracing drops to the ridge immediately beneath and continues from there.

Classification. Although the system used by the Federal Bureau of Investigation is based on the system devised by Sir Edward Henry, it has been necessary to amplify and extend the Henry system with numerous additional subdivisions. The 10 fingers are considered as a unit in obtaining the complete classification. This classification is expressed in a formula consisting of a combination of letters and numerals. It reflects the general characteristics of the patterns in all fingers and permits filing in proper sequence for ready reference.

The fingerprint pattern types are indicated at the bottom of each finger block. Under the index fingers the appropriate capital letters are placed for every pattern except the ulnar loop, which is designated by a diagonal line slanting in the direction of the loop. Whorls in any finger are designated by the letter W. Under all other fingers, the appropriate small letter or diagonal line is placed for every pattern.

Formula. The classification formula is generally composed of the following divisions: primary, secondary, small-letter group, subsecondary, major, final, and key.

For the purpose of obtaining the primary classification, numerical values are arbitrarily assigned to each of the 10 finger blocks. Fingers numbers 1 and 2 have a value of 16, fingers numbers 3 and 4 have a value of 8, fingers numbers 5 and 6 have a value of 4, fingers numbers 7 and 8 have a value of 2, and fingers numbers 9 and 10 have a value of 1. Wherever a whorl appears it assumes the value of the space in which it is found. The summation of the numerical values of the whorl type pattern, if any, appearing in fingers 1, 3, 5, 7, 9, plus an arbitrary 1, furnishes the denominator of the primary. The summation of the values of the whorls, if any, in fingers 2, 4, 6, 8, 10, plus an arbitrary 1, furnishes the numerator of the primary. For example, the primary classification on the set of fingerprints (Fig. 4) is derived from the location of the whorls in fingers 7, 9, and 10. The numerator is based on the whorl in the tenth finger which has a value of 1. This plus the 1 arbitrarily added gives the 2 in the formula. The denominator depends upon the values of 2 and 1 given the whorls in fingers 7 and 9. This total of 3 plus the arbitrary 1 yields the 4 shown in the denominator.

It should be noted that after the primary is obtained the entire remaining portion of the classification formula is based upon the arrangements of the impressions appearing in the right hand as the numerator over the impressions appearing in the left hand as the denominator.

Secondary. The secondary classification appears to the right of the fractional numerals which represent the primary. It is shown in the formula by capital letters representing the types of patterns appearing in the index fingers of each hand.

Small letters. The small-letter group includes the prints with an arch or tented arch in any finger or a radial loop in any except the index fingers. Such small letters are brought up into the classification formula in their proper relative positions, either to the left or right of the secondary. A dash is used to indicate the absence of each small letter between the index fingers and another small letter or between two small letters.

Subsecondary. The subsecondary classification is obtained by grouping the ridge counts of loops and the ridge tracings of whorls. The ridge counts are translated into small and large, represented by symbols I and O. The whorl tracings are brought up as I, M, or O, denoting inner, meeting, or outer ridge tracings of the whorl types. Only six fingers are used to obtain the subsecondary, numbers 2, 3, 4, in the numerator, and 7, 8, and 9, in the denominator. The subsecondary is placed to the right of the secondary in the classification formula.

Major. The major classification is derived from the thumbs. When whorls appear, the major reflects the whorl tracings. When loops appear, however, a table of ridge count values is used to translate the ridge counts into small, medium, or large groups, designated by the letters S, M, L.

Fig. 4. Set of fingerprints showing classification and pattern types. *(Police Department, Charlottesville, Va.)*

major is placed to the *left of* the primary, in the classification formula.

Final. The final provides a definite order of filing the prints within the aforementioned subdivided groups. It is the ridge count of the loop in the right little finger. If the right little finger is not a loop, the ridge count of the left little finger is used. It is indicated at the extreme right of the classification formula.

Key. The key is obtained by counting the ridges of the first loop appearing on the fingerprint card, beginning with the right thumb, exclusive of the little fingers which are never considered for the key because they are reserved for the final. The key is always placed to the extreme left of the numer-

... in corresponding fingers. This enables the location of a record within minutes by examining a limited number of the millions of individual cards on file.

Latent prints. The latent fingerprint section of the Federal Bureau of Investigation deals with the identification of single or latent (hidden) fingerprints developed at the scene of a crime or upon articles of evidence. This generally involves the examination of fragmentary latent finger, palm, or even foot impressions developed by appropriate processes on objects associated with various crimes.

The distinctive ridges of the palms of the hands and the soles of the feet closely resemble those of the fingerprint. The fingerprint system is, therefore, easily applicable to the classifying of palm or sole impressions. Although palm or sole prints are not utilized as much as fingerprints, they are accepted as a positive assistance to law enforcement officers when fingerprints are not available.

[J E H]

Fiord

A narrow, deep, steep-walled marine embayment found on glaciated coasts, mostly mountainous, particularly in Alaska, Labrador, Greenland, Norway, and southernmost South America. The depth of water is generally least at the entrance from the ocean; a maximum of 4000 ft has been recorded. Most fiords are perpendicular to the coast, but some are parallel to it. The pattern of development is either dendritic or rectangular (trellis). Many tributary valleys show hanging junctions, some of them submerged. Some fiords have islands, and the depth of water on the longitudinal axis is very irregular. Fiords have been ascribed to sinking of a stream-eroded coast to form drowned valleys, to faulting, and to glacial erosion. Since all fiords occur in glaciated regions, it is clear that glaciers must have had some part in their formation. Localization of erosion is related to preglacial valleys which molded the ice into narrow tongues. Although the sides of many fiords are glacially striated and polished, only a portion of the rock which was removed could have been ground into powder. Plucking of large masses of rock beneath thick ice is a much more rapid process of erosion, for it requires far less energy. This process is possible only where the glacier is thick enough to permit pressure melting and refreezing around rock masses which were previously fractured by earth movement. *See* COASTAL LANDFORMS.

The rectangular pattern of many systems of fiords is related to regional structure, including faulting, and thus ties in the tectonic theory of fault control with the hypothesis of glacial erosion. Fiord islands are localities of minimum fracturing and erosion. Change in level of land and sea is not important in this theory. Coasts which show a fiordlike plan on the map, such as those of Maine and the northeast Adriatic, have shallow bays unlike true fiords. Although the formation of fiords requires narrow tongues of glacial ice, the mountainous topography of the land beneath the glaciers of Greenland and Norway lent itself to the requisite concentration of ice flow. Study by F. E. Matthes of Yosemite Valley, California, proved the importance of glacial plucking of fractured rock.



Fig. 1. Fiord in southeastern Alaska. (W. W. Atwood, USGS)

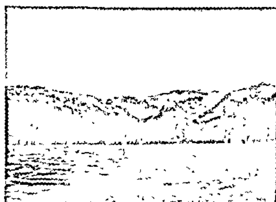


Fig. 2. Typical fiord-side scenery of southeastern Alaska. Steep forested slopes give way upward to rocky ice-scoured upland in this June view. (R. M. Chapman, USGS)

High waterfalls from hanging tributaries tell of widening and deepening of a preglacial valley by this process. The several different levels of hanging valleys show that the region was tilted in several distinct stages, thus accelerating erosion of down-slope streams and the origin of discordant stream junctions not related to glaciation. Many deep inland lakes, including the Finger Lakes of New York, are termed inland fiords. *See* FINGER LAKES; GLACIATED TERRANE. For a treatment of water circulation in fiords (marginal seas) *see* OCEAN CURRENTS. [F.T.T.]

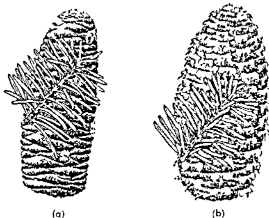
Bibliography: J. W. Gregory, *The Nature and Origin of Fiords*, 1913; F. E. Matthes, *Geologic History of the Yosemite Valley*, USGS Profess. Paper 160, 1930; F. T. Thwaites, *Outline of Glacial Geology*, 1956.

Fir

Any tree of the genus Abies of the pine family, characterized by erect cones, by the absence of resin canals in the wood but with many in the bark, and by flattened needlelike leaves which lack definite stalks. The leaves usually have two white lines on the under side and leave a circular scar when they fall. The native fir of the northeastern United States and adjacent Canada is *A. balsamea* which attains a height of 75 ft and has resinous buds. Its principal uses are for paper pulp, lumber, boxes and crates and as a source of the liquid resin called Canada balsam. In the eastern United States, the fir is commonly used as a Christmas tree. It does not do well as an ornamental tree where the summers are hot.

The Fraser fir, *A. fraseri*, is a similar species found in the southern Appalachians. In this species the bracts grow out beyond the cone scales.

Several species of *Abies* grow in the Rocky Mountain region and westward to the Pacific Coast. The most important is the white fir, *A. concolor*, also known as silver fir, a large tree which may reach a height of 120 ft and is characterized by blue-green needles 2–3 in. long. It is often



Cones and branches of (a) Balsam fir, *Abies balsamea*; (b) Fraser fir, *A. fraseri* (USDA)

vated as an ornamental in the eastern United States, and also furnishes lumber for construction timber, boxes, crates, and general millwork.

Other western species of commercial importance are the subalpine fir, *A. lasiocarpa*; grand fir, *A. grandis*, Pacific silver fir, *A. amabilis*; California red fir, *A. magnifica*; and the noble fir, *A. procera*. The lumber cut from the western true firs now totals more than 1,000,000,000 board ft annually, with California furnishing more than one-half. A large part of this is white fir. *See* FOREST AND FORESTRY; TREE [A.H.G.]

Fire

A rapid but persistent chemical reaction accompanied by the emission of light and heat. The reaction is self-sustaining, unless extinguished, to the extent that it continues until the fuel concentration falls below a minimum value. Most commonly, it results from a rapid exothermic combination with oxygen by some combustible material. Flame and heat also may result from a reaction involving an agent other than oxygen. Thus, certain reactive metals such as zinc will burn in an atmosphere of chlorine.

Flame, the visible manifestation of fire, results from a heating to incandescence of minute particulate matter composed principally of incompletely burned fuel. The color of the flame depends upon the material undergoing reaction and the temperature. *See* COMBUSTION; FLAME; FUEL. [F.J.J.]

Fire detector

A temperature-sensing device designed to sound an alarm, to turn on a sprinkler system, or to activate some other fire preventive measure at the first signs of fire. Sprinkler systems are usually equipped with fusible metal links, and these are perhaps the most common fire detector devices. They really serve a double function because they sense the rise in temperature due to the fire and put the defense measures into motion.

More elaborate fire alarm systems use detectors similar to thermostats. They are often actuated electrically. Although some types become actuated simply when the temperature rises above a minimum set value, others can be set to operate when the rate-of-rise of temperature exceeds a certain value. Still more complex systems involve effects due to a combination of rate-of-rise and temperature level.

Smoke detectors are also common in certain special applications such as fur storage warehouses and airplane baggage compartments. These detectors are actuated when the smoke interrupts a light beam that shines on a photoelectric cell. Maintenance of this equipment in reliable working order is a more serious problem than it is with most other detectors.

Photosensing devices are now coming into use that respond to a stimulus in the infrared part of the spectrum and can thus detect the rise in temperature at the source of the fire itself. Because of the rapid response of such detectors, extinguishing apparatus can be actuated in a fraction of a second. Therefore, it is possible to stop the development of an explosion in the fuel tank of an aircraft, for example, by placing the detector and the extinguishing equipment inside the tank itself. *See* ALARM SYSTEMS; FIRE EXTINGUISHER. [W.E.G.]

Fire extinguisher

Fire may be extinguished by the following methods: (1) cooling the burning materials; (2) blanketing the fire with inert gas that chokes it for lack of oxygen; (3) introducing materials that inhibit combustion; and (4) covering the burning matter with a blanket or a layer of solid particles that prevent access of air. Fire extinguishers operate on one or a combination of these principles.

The method or methods that are used to put out a fire depend upon its nature. Burning magnesium reacts with water according to the equation: $Mg + H_2O \rightarrow Mg + H_2$. This reaction is itself a form of combustion which is supported by water instead of oxygen. In fact, the hydrogen evolved will burn in the air and possibly create a worse situation. However, to pursue this example still further, the action of water on a magnesium fire will depend not only on these chemical factors but also on the relation between the amount of water and the amount of burning magnesium (the surface area of magnesium covered by flame). A small amount of water on a large magnesium fire would make matters worse; but a very large amount of water on a small magnesium fire would quench it because of the cooling effect. Therefore, both the chemical and the physical nature of the fuel, and the size of the fire have an important bearing on the effectiveness of a given extinguishing technique.

Water. Water is the most effective cooling agent used in fire extinguishing. The generation of steam also drives away the air and forms a blanket, but being less dense than air, it is rapidly displaced.

Wetting agents and foaming agents increase the effectiveness of water. In the case of burning cotton

bales, for example, the addition of a wetting agent will cause the water to penetrate the burning material and extinguish the fire more efficiently. Foaming agents make it possible to produce a heavy blanket of foam over the fire. This excludes air and cools the burning material at the same time. Attachments that feed wetting and foaming agents into the fire hose stream are now common.

The use of water in the form of fog rather than as a spray or stream is very effective under some conditions. In fighting an oil fire, for example, fog is effective in cooling the fire and providing a blanket, whereas a stream of water would serve only to disperse the oil and spread the fire. The greatest benefit of fog is that it provides a protecting screen and allows the fireman to work in close to the fire.

In the small first-aid water fire extinguishers, a propellant must be provided. Usually this is carbon dioxide, which is either generated when needed (the soda-acid extinguisher) or stored in a cartridge. When a foaming agent is used in such an extinguisher, the foam is rendered still more effective by the inert gas that fills the bubbles.

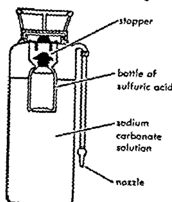
Antifreeze agents must be added if the water extinguisher is placed outdoors in cold climates. Commonly, calcium chloride is used, but in certain cases (the so-called loaded stream extinguishers), the antifreeze salts may be such as to have a definite fire retarding effect.

There are many fires, as pointed out above, on which water should not be used. On oil fires, for example, water is effective only as a foam, or in some cases as a fog; a water stream should never be used. On electrical fires, the use of water can prove extremely dangerous, and in the case of electric generators and motors, it is completely ruinous. Fog nozzles can sometimes be used effectively, but there is always an element of added danger to the fireman whenever water is used near an electrical fire. Carbon dioxide or dry extinguishers are decidedly preferred.

Automatic water sprinkler systems are now the most common form of fire protection in industrial plants and large buildings. The installation of these systems is the greatest single factor to be credited for the sharply reduced incidence of disastrous fires in recent years.

Carbon dioxide. A safe and effective extinguisher for all confined fires, carbon dioxide, CO_2, acts as an inert blanket, and because it is heavier than air, it will exclude oxygen very efficiently from a fire on the floor of a building, or in a vat or similar vessel. It is not effective on a fire in an elevated location, or outdoors where the wind can blow it away.

Carbon dioxide is supplied as *liquid in a tank.* The pressure in the tank is about 800–900 psi, small leaks are therefore serious, and the tank should be weighed frequently to ensure that the contents are still intact. Because of the high pressure, large carbon dioxide containers are not practical. For small hand-carried extinguishers, the carbon dioxide tank has a short hose with a horn



Soda-acid fire extinguisher. When the extinguisher is inverted, the stopper falls out of place, allowing the sulfuric acid to react with the sodium carbonate to produce carbon dioxide.

at the end through which the gas may be directed onto the fire. Large installations involve a battery of carbon dioxide cylinders attached to a manifold line.

In certain industrial applications, distribution systems are used for carbon dioxide analogous to the sprinkler systems for water. For example, when a fire starts in electric power generating stations, jets are opened to blow CO_2 into the generator housings. This quickly blankets the critical area with the gas. Similar equipment is provided in large spray-painting operations.

In some industrial plants, provision is made to use flue gases to extinguish fires. The flue gas must be cooled and washed before it is put on the fire. Steam from a boiler is similarly used in certain special instances, but its application is definitely limited. *See* CARBON DIOXIDE

Dry chemical extinguishers. A dry powder, consisting principally of sodium bicarbonate, is gaining wide use as a fire extinguisher. The powder must have the correct particle size and contain materials that prevent it from caking. The action of the powder is threefold. Probably its most important action is to generate carbon dioxide by the reaction, $2NaHCO_3 \rightarrow Na_2CO_3 + H_2O + CO_2$. Because this reaction is caused by the heat of the fire, the carbon dioxide is generated precisely where it is needed. Secondly, the powder serves to cool the burning material, and thirdly, it provides a shielding to prevent access of air. Dry chemical is useful for small fires, and especially electrical fires. As added advantages, sodium bicarbonate is nontoxic, does little or no secondary damage, and is easily cleaned up after the fire.

Halogenated hydrocarbons. Carbon tetrachloride, CCl_4, has had a long history as a fire extinguishing agent. As it is customarily used, in small quantities, the principal action is to supply a heavy blanket of vapor over the fire. Because it is 3.5 times as dense as carbon dioxide, it is very effective for this purpose.

In addition, carbon tetrachloride, in common with all the halogenated compounds, has a definite chemical inhibiting effect on combustion. Even vapor percentages in the air as low as 4.5% will sometimes put out a fire. Larger percentages are usually needed; but the halogenated materials are more effective by far in suppressing the inflammability of gasoline vapor, for example, than inert gases such as carbon dioxide and nitrogen.

Other halogenated hydrocarbons that have been used besides CCl_4 are chlorobromomethane, CH_2ClBr, and several of the fluorinated hydrocarbons commonly known as freons. An effective agent that has recently become of interest is Freon 13B-1, CF_3Br.

The principal difficulty, however, with all the halogenated hydrocarbons, and with CCl_4 in particular, is toxicity. The vapors of carbon tetrachloride are themselves quite dangerous. They may [illegible]

addition to other noxious fumes, small percentages of phosgene, $COCl_2$, may be generated, and this gas is most deadly. Certain of the freons are much less toxic than CCl_4 and less apt to give dangerous products. Interest is therefore centering in these new materials. *See* HALOGENATED HYDROCARBON.

Other extinguishing methods. Small fires are best treated by choking off the supply of air. In addition to the several methods for doing this already described, others should be mentioned that require no special equipment. For a household fire, especially involving clothing on a person, a blanket or a rug provides an effective means to smother the fire. Small fires around a kitchen stove may be snuffed out with salt, or, better still, with bicarbonate of soda. A bucket of sand, strategically located, is also useful against domestic fire hazards.

A most novel technique for putting out a fire [illegible]

vided for doing this.) As a result of the agitation, the cold oil from the bottom is brought up to displace the hot burning layer of oil at the top. This method, where applicable, is amazingly effective.

Still other firefighting methods are called for in various special cases. One of these is the explosion of firedamp (methane) in mines. The explosion generates a wind in the mine shaft because the air is pushed ahead of the flame. This accelerates the progress of the explosion and increases the devastation. It has been found that stone dust will stop such fires when it is spread along the mine passages so as to be stirred up readily by the wind. The stone dust probably cools the hot gases to the point where the flame will not propagate.

In the case of forest fires, different techniques, such as back-firing, are used. These depend on wind conditions, moisture, and fuel type. Water suspensions of sodium calcium borate have proved effective in preparing a fire break and even in direct attack on the fire. Borates are well known as fire retardants for paper and fabrics. *See* COMBUSTION; EXPLOSION AND EXPLOSIVE; FIRE DETECTOR; FLAME; FLAMEPROOFING. [W.E.C.]

Bibliography: National Fire Protection Association, *Handbook of Fire Protection*, 11th ed., 1954; National Safety Council, *Accident Prevention Manual for Industrial Operations*, 3d ed., 1955.

Fire-control systems

Equipment for controlling the aiming and firing of [illegible] batteries or missile launchers. Fire-control sys[illegible] artillery, [illegible]s, and the [illegible]. *See* CON[illegible]

The basic components of a fire-control system (Fig. 1) are (1) sensing elements, (2) a computer, and (3) gun-positioning servos. Figures 2 and 3 show two diverse fire-control systems. *See* COMPUTER; INSTRUMENTATION; SERVOMECHANISM.

In the manual system of Fig. 2 the sensor is the man's eye, which he keeps on the target and with which he judges the direction and distance to the target and its velocity relative to his gun barrel. Using this information the man's brain (computer) estimates the amount by which the gun barrel should lead the bird. Finally, the man uses his arm muscles (servo) to bring the gun barrel to the "computed" lead angle and to move the barrel at the correct rate to maintain that angle. Corrections are made continually for changes in target velocity, for vehicle maneuvers, and other factors that affect the geometrical problem. Some fire-control terminology is labeled in Fig. 2.

Antiaircraft system. In Fig. 3 the sensor is a radar, mounted on a gun director to measure direction and distance to the target aircraft. An automatic tracking loop keeps the antenna aimed at the moving target (just as the man kept his eye on the target in Fig. 2).

By noting the rate of rotation of the antenna and the range to the target, the computing equipment (in the director) calculates the correct lead angle. Specifically, if T is projectile time-of-flight, and if the antenna is rotating at rate Ω, then the line-of-sight will rotate through an angle ΩT between firing and impact. Thus, to a first approximation, ΩT is the lead angle.

Time-of-flight T depends upon range R, or distance to the target, and upon missile velocity V.

$$T = R/V$$

The gun servos are powerful, fast-acting, hydraulic devices. Two are required for each gun—one for azimuth (rotation about a vertical axis) and one for elevation.

Computing optical sight. Before completely automatic tracking radars were available, antiaircraft fire was accurately controlled by the semiautomatic computing optical sight. This device uses a series of mirrors to present the operator with an optical image, known as a reticle. The mirrors are

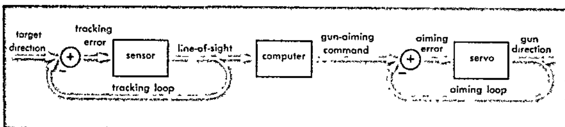


Fig 1. Elements of fire-control system.

mechanically linked to rate gyroscopes (which measure the turning rate of the case, or sight head) so that the reticle is displaced from the line of sight by an angle proportional to the sight-head rate. Thus, if the operator moves the sight so as to keep the reticle superimposed upon the target, the sight head will automatically be pointing at the correct lead angle. *See* GYROSCOPE.

In early systems range was estimated by an assistant who set a dial on the sight head to adjust the mirror linkages appropriately. Later, range radar was added.

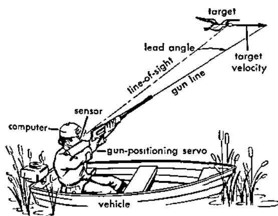


Fig 2. Manual "fire-control system."

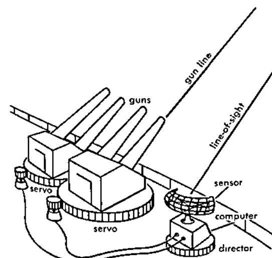


Fig 3 Automatic antiaircraft fire-control system.

Naval artillery systems. Large naval guns (5–16 in.) and rocket launchers used against shore batteries and ships at long range are controlled by slower, more precise, large scale versions of the equipment shown in Fig. 3. The director carries heavy radar and telescope equipment, operated by several men. The elaborate computer is also operated by a team of men who make ballistic and miss corrections.

Bomber defense systems. Bomber defense systems, which direct the gun turrets in bomber aircraft, are much smaller and faster-acting than the equipment in Fig. 3 An entire system of several guns may be operated by one man who uses a computing optical sight to aim his own ball-shaped turret at the target. The other gun turrets follow by servo control.

Interceptor aircraft systems. When rockets are launched from a fighter airplane the entire airplane becomes the "gun barrel" and must be aimed by the fire-control system on board. Superimposed images, computed from a tracking radar or from lead-computing optical sights are given the pilot. He may use them to fly the plane himself, or they may be electrical steering signals to the automatic pilot to effect completely automatic interception. *See* AUTOPILOT

The computer must be light and compact and must solve a complex set of vector equations to initiate the mode of attack selected by the pilot

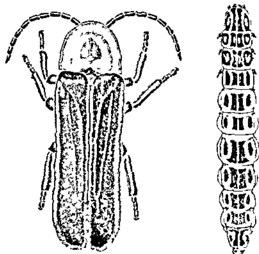
When a radar signal is used to control automatically the flight of a fighter airplane, the airplane may display violent motions in its attempt to follow target glint noise caused by random wandering of the radar beam as it tracks the apparent center of the maneuvering target (*see* MISSILE GUIDANCE SYSTEMS). Much of this noise may be attenuated by electrical filtering networks, but only at the expense of reducing the speed of system response, so that a compromise must be made to maximize kill probability

Broadly, ground command stations which direct an entire flight of interceptor aircraft or guided or ballistic missiles are fire-control systems. Perhaps the ultimate is the SAGE system whose several huge digital computer installations direct the operations of an entire continental defense arsenal. [R.H.C.]

Bibliography: T. A. Morgan, Navy's gyro gun sight, *Army Ordnance*, 29(151):68, 1945; W. Wrigley and J. Hovorka, *Fire Control Principles*, 1959

Firefly

Any of several light-producing insects of the family Lampyridae, order Coleoptera. Most, but not all, of the insects in this family produce light. There are about 2000 species, primarily tropical and subtropical, about 20 of which occur in the United States. These are soft-bodied, elongated, nocturnal



The firefly, *Photinus pyralis*; length ½–¾ in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

beetles of medium size, somewhat flattened, with dark outer wings, usually black, brown, or yellow. The light they produce is usually yellow, but may be green, orange, or red. In some species only females are luminescent; in others both sexes may produce light. The wingless females and larvae of some species, when luminescent, are known as glowworms. See BIOLUMINESCENCE; COLEOPTERA. [J.D.B.]

Fire-tube boiler

Hot products of combustion pass through tubes which are surrounded by the boiler water. Such boilers require an outer shell for containment of water and steam at elevated pressure. For this reason they are often referred to as shell-type boilers. The principal heat-transferring surface is provided by tubes or flues of relatively small diameter, connected to tube sheets at each end of a cylindrical shell, through which hot gases flow. These tubes serve also as structural reinforcement, to support the flat tube sheets against the force of internal pressure. Braces or tension rods are used for areas of the sheet not covered by the tubes.

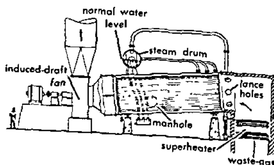


Fig. 1. Horizontal-return fire-tube boiler.

Fire-tube boilers may be designed for vertical, inclined, or horizontal position. One type that has been widely used is the HRT or horizontal return tubular boiler (Fig. 1) in which part of the heat is delivered directly from the furnace to the lower portion of the shell. The gases make a return pass through the flues before being discharged to the stack. In the Scotch marine boiler, one or more large flues provide a combustion space within the shell, and gas is returned through groups of smaller flues. The large flues are corrugated to resist collapsing pressure. These boilers may be fired by oil burners or by solid fuel burned on grates set within the furnace flues.

The conventional locomotive boiler represents a combination of water-jacketed firebox and fire-tube boiler. Staybolts or links are connected between the inner and outer shells of the firebox for reinforcement against internal pressure. Superheating of the steam is accomplished by tubular elements positioned within the flues.

Gas-tube boilers are sometimes used for absorption of waste heat from process gases or exhaust from internal combustion engines, where adaptation to the duct provides a simple and economical installation for recovery of low-grade heat. The shell may be augmented by an external steam separating drum, with downcomer and riser connections to maintain active circulation (Fig. 2).

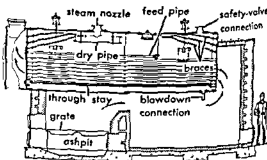


Fig. 2. Fire-tube boiler for recovery of heat from waste process gas.

Shell-type boilers are limited in size and operating pressure by the allowable stresses set up in large-diameter shells and flat tube sheets of practicable thickness. They are best suited for low-pressure heating service and for portable use. Inherent advantages are the integrated structure that can be readily transported, and the generous reserve water capacity requiring minimum attendance. Inherent disadvantage lies in the risk of catastrophic damage if overheating and rupture of the shell or tube sheet occurs as a result of low water or the formation of an insulating layer of internal scale caused by water impurities. See STEAM GENERATING UNIT. [F.G.E.]

First-order transition

A change in state of aggregation of a system accompanied by a discontinuous change in enthalpy, entropy, and volume at a single temperature and pressure. This transition may be between liquid and gas (see BOILING POINT), between solid and gas (see SUBLIMATION), between solid and liquid (see MELTING POINT), or between two solid phases (see TRANSITION POINT). For the differences between first- and second-order transitions, see SECOND-ORDER TRANSITION. See also EQUILIBRIUM, PHASE. [R.L.S.]

Fischer-Tropsch process

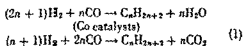
The synthesis of gaseous, liquid, and solid hydrocarbons and, to a lesser extent, of aliphatic alcohols, aldehydes, and ketones by the catalytic hydrogenation of carbon monoxide. The process was discovered in 1923 by F. Fischer and H. Tropsch at the Kaiser Wilhelm Institute for Coal Research in Mülheim-Ruhr, Germany.

Catalysts. In the period 1923–1935, F. Fischer and his associates developed highly effective catalysts and measured the effects of different temperatures, pressures, and hydrogen to carbon monoxide ratios on the yield and composition of the product. The first industrial plant was erected in 1935 at Oberhausen, Germany. In 1938, when construction of Fischer-Tropsch plants and engineering development of the process were halted in favor of additional coal hydrogenation plants, six German plants for hydrogenation of carbon monoxide were producing primary products at the rate of about 4,000,000 bbl/yr. The synthesis gas, prepared in standard water-gas plants, contained 2 volumes of hydrogen per volume of carbon monoxide. The catalyst was a fixed bed of granules and contained in parts by weight: 100 cobalt, 5 thoria, 8 magnesia, 200 diatomaceous earth. The synthesis was conducted at 200°C (392°F) and at pressures of 1–10 atmospheres (atm). The average yield per cubic meter of synthesis gas was 150 grams of hydrocarbons ranging from propane-propylene to waxes of 2000 molecular weight and including minor amounts of oxygen-containing organic compounds. The hydrocarbon product distribution in per cent was 10 with 3 and 4 carbon atoms, 40–50 gasoline, 15–25 diesel oil, 10–20 soft wax (boiling range 280–340°C, or 536–644°F), and 5–15 hard wax.

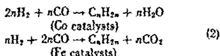
Because of the limited supply of cobalt in Germany during World War II, iron catalysts of sufficient activity were developed. Iron catalysts utilize hydrogen and carbon monoxide in the ratio of 0.8–1.0, making possible the use of water-gas directly, whereas for cobalt catalysts the hydrogen content must be increased by further reaction with steam. Cheaper and more durable iron catalysts were developed in the United States using mill scale or magnetite ore to which about 0.5% of potassium (as K_2CO_3) was added. These catalysts, when pretreated with carbon monoxide, fo[rm] carbides which are active at pressures [...] atm and temperatures in the range 250–325°C (482–617°F). At 300–320°C the yield of hydrocarbons is about the same as given above for cobalt catalysts. However, about 70% of the hydrocarbons of the product boil in the gasoline range, and the octane rating is much higher than that of the gasoline fraction of the product from cobalt catalysts. The amount of oxygen-containing organic compounds is about 15% of the total product from iron catalysts as compared with less than 5% of that from cobalt catalysts.

Two industrial-scale plants using iron catalysts (mill scale plus potash) have been erected, one at Brownsville, Texas, and another in South Africa. Both plants use finely divided catalyst suspended in the gaseous reactant, the plant in Brownsville having a conventional "fluidized" bed and that in South Africa an entrained catalyst system with recycle of catalyst. Operation of both plants has presented engineering difficulties which have not been completely overcome. Also, in South Africa a full-scale plant using a fixed-bed granular catalyst is in operation. This plant uses a catalyst prepared by precipitation of a mixture of ferric and ferrous iron hydroxides and cupric hydroxides from aqueous solution of the nitrates. The finished catalyst contains also about 0.5% of potassium.

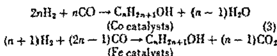
Reactions. Typical Fischer-Tropsch reactions for the synthesis of paraffins (1), olefins (2), and alcohols (3) are:



$$\begin{aligned}(2n+1)H_2 + nCO &\rightarrow C_nH_{2n+2} + nH_2O \quad \text{(Co catalysts)}\\ (n+1)H_2 + 2nCO &\rightarrow C_nH_{2n+2} + nCO_2 \quad \text{(Fe catalysts)}\end{aligned} \tag{1}$$



$$\begin{aligned}2nH_2 + nCO &\rightarrow C_nH_{2n} + nH_2O \quad \text{(Co catalysts)}\\ nH_2 + 2nCO &\rightarrow C_nH_{2n} + nCO_2 \quad \text{(Fe catalysts)}\end{aligned} \tag{2}$$



$$\begin{aligned}2nH_2 + nCO &\rightarrow C_nH_{2n+1}OH + (n-1)H_2O \quad \text{(Co catalysts)}\\ (n+1)H_2 + (2n-1)CO &\rightarrow C_nH_{2n+1}OH + (n-1)CO_2 \quad \text{(Fe catalysts)}\end{aligned} \tag{3}$$

The primary reaction on both cobalt and iron catalysts yields steam, which reacts further on iron catalysts with carbon monoxide to give hydrogen and carbon dioxide. On cobalt catalysts, at synthesis temperatures (about 200°C, or 392°F), the reaction $H_2O + CO = CO_2 + H_2$ is much slower than on iron catalysts at synthesis temperatures (250° − 320°C, or 482° to 617°F). All of the above synthesis reactions are exothermic, 37–51 kilocalories per mole of carbon in the products or 4700–6100 Btu/lb of product. The heat evolved must be rapidly dissipated to avoid temperature increases which would result in extensive production of methane and rapid deterioration of the catalyst.

The hydrocarbons from iron catalysts contain more olefins than those from cobalt catalysts. The products from both catalysts are largely straight-chain aliphatics; the per cent of branching is abou[t] 10 for C_4, 19 for C_5, 21 for C_6, and 34 for [...] Aromatics appear in small amounts in the

in larger amounts in the higher boiling fractions. Operating conditions and special catalysts are known for the production of higher proportions of alcohols. When olefins are added to the synthesis gas and the reaction is conducted at pressures of about 200 atm (3000 psi) on cobalt catalysts, the product is largely an aldehyde with one more carbon atom than the starting olefin. This addition of hydrogen and carbon monoxide to olefins is known as the Oxo process and is extensively used for the synthesis of specific aldehydes and alcohols from olefins. *See* BERGIUS PROCESS; HYDROFORMYLATION. [H.H.ST.]

Bibliography: H. Pichler, Twenty-five years of synthesis of gasoline by catalytic conversion of carbon monoxide and hydrogen, *Advances in Catalysis*, 4:271–341, 1952; H. H. Storch, N. Golumbic, and R. B. Anderson, *The Fischer-Tropsch and Related Syntheses*, 1951.

Fish

In common usage, any aquatic vertebrate with fins instead of legs. Some authorities do not include in their definition of the term fish the lampreys and hags, which lack paired fins and have a very primitive vertebral column.

Fishes first appeared in the Ordovician geologic period and were the only vertebrates until the Late Devonian, when they gave rise to the amphibians. There are about 30,000 living species, all but about 300 of which belong to the Osteichthyes or true bony fishes

Economically, fishes are the most important of all wildlife Fish has been a significant item in the diet of man since the earliest times. Today the fisheries of the world harvest about 35,000,000,000 lb annually with a gross value of over $800,000,000. However, the utilization of fish is by no means limited to food; fish oil, meal, scales, and other byproducts serve a wide range of uses in agriculture, medicine, and industry. Fishing is the leading sport of the United States in number of participants and in total expenditure. The 18,000,000 sports fishermen in the United States spend in excess of $2,000,000,000 annually on fishing.

[illegible]

[illegible]. Fish farming, including the rearing of fishes for bait, by hobbyists, and for food, has grown rapidly in the past few years, and appears destined to assume still more prominence in the future. *See* PISCES (ZOOLOGY). [J.D.B.]

Fisher

A large carnivore, *Martes pennanti*, of the family Mustelidae, which is confined to the dense northern and western mountain forests of North America. [illegible]

[illegible] martens and squirrels, and it is one of



The fisher, *Martes pennanti*; length 3 ft. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

the few animals that prey upon the porcupine. Its fur is valuable, and it is easily trapped. As a result, it has become quite rare in the southern part of its range. *See* CARNIVORA. [J.D.B.]

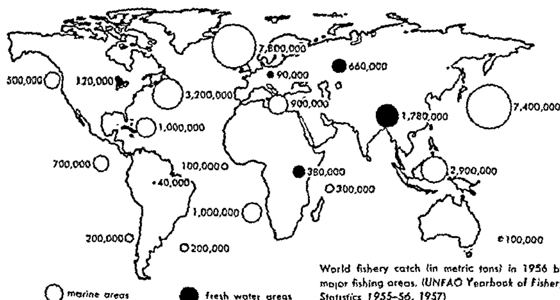
Fisheries conservation

The term fishery is used here to mean the place for taking fish or other aquatic life, particularly in sea waters. The term has various legal connotations depending upon whether the right to take fish is founded on ownership of the underlying soil and therefore exclusive (several fishery), whether the right to fish in public waters is enjoyed in common with others (common fishery), or whether the right is an exclusive privilege, derived from royal or public grant, and independent of the ownership of the underlying soil (free fishery).

A fishery may be operated for pleasure (sport fishery) or for profit (commercial fishery). Fisheries include finny fish; mollusks, such as clams, mussels, and oysters; shellfish, including lobsters, shrimp, and crayfish; aquatic plants; sponges; coral; sea cucumbers; amphibians, chiefly frogs; reptiles such as turtles, alligators, and crocodiles; and mammals including whales, seals, and walruses (*see* MARINE FISHERIES; MARINE RESOURCES).

Fisheries are important for the production of food and of raw materials for industry, and for recreation. They occur in all the many kinds of waters that together compose nearly three-fourths of the earth's surface (Fig. 1). The two principal types, however, are marine (salt-water) fisheries, which annually yield 25,000,000–30,000,000 metric tons of products; and fresh-water (inland, continental), producing annually a recorded catch of 3,000,000–4,000,000 metric tons plus a sporting take of great but unknown magnitude. Both types have potentials far greater than present realizations.

Marine fisheries are mostly commercial and are located predominantly in the Northern Hemisphere The most important fishery centers and the principal products they yield are the northeastern Atlantic (flatfishes including halibut and flounder, cod, haddock, coalfish, herring, and shrimp); the northwestern Pacific (salmon, flatfish, herring, crab, shrimp, lobster, squid, and octopus); the



World fishery catch (in metric tons) in 1956 by major fishing areas. (UNFAO Yearbook of Fishery Statistics 1955–56, 1957)

northwestern Atlantic (flatfish, cod, haddock, herring, lobster, and crab); and the Indo-Pacific (herring, bonito, mackerel, and shrimp).

The major fresh-water fishing areas and their products are Asia (ayu, salmon, milkfish, carp); Russia, (salmon, whitefish); Africa (many kinds of fish); and central and northern North America (trout, whitefish, bass, perch and their relatives) Fresh-water fisheries in the more highly civilized parts of the world are used primarily for sport. However, they are important largely for food in primitive regions including parts of Africa and South America, and in densely populated lands, especially those on a rice economy, such as India, China, Japan, and the smaller nations of southeast Asia.

Fisheries conservation problems. The major problem in fishery conservation is how to control both man and the aquatic community to insure the production and the harvest of aquatic crops for the present and for the future when the demand will probably be greater than now. Lack of knowledge or care by fishermen may make their fishing efforts inefficient, their treatment of captured fish wasteful, and their methods of capture destructive to stocks needed as "seed" for future harvests Handlers, marketers, and consumers may cause waste by careless preservation (refrigeration, salting, and canning), and by inefficient preparation for the table. Lack of information may also be responsible for inadequate harvest of many underexploited or unexploited segments of the resource. Consumers need to be educated away from highly preferential buying habits that accelerate the demand and price for certain species but lead to the discard of other kinds of aquatic organisms with equally sound food values.

The destruction of the aquatic habitat by man accentuates fishery problems, especially in inland waters. Deforestation and destructive agriculture, resulting in soil erosion and excessive warming of waters, have changed the fish producing characteristics of many streams, usually for the worse (*see* FOREST CONSERVATION; SOIL CONSERVATION). Sewage and industrial wastes reduce the quality of water suited for desired aquatic life (*see* WATER POLLUTION). Organic pollutants such as domestic sewage may remove life-important oxygen from the water (waters generally are undesirable for aquatic life if the dissolved oxygen in them falls below 3 ppm) Pollutants may also be directly poisonous to fish and other water organisms (for example, most chemicals that are poisonous to man, such as cyanides, are also toxic to desirable aquatic life). Habitats may be destroyed by smothering them with silt (washings from mineral refineries such as coal and iron).

Man has likewise created problems for aquatic life in continental waters by changing phases of the water cycles. Deforestation and agricultural land drainage have lowered the ground water table and have lessened stream flows during dry periods. The construction of dams for water storage, power, navigation, industry, flood control, domestic water supply, and irrigation has interfered with the movements (usually for spawning) of native migratory fish. Outstanding problems created by dams are those affecting the migrations of the Columbia River salmon in the western United States and the Atlantic salmon throughout its range in both northeastern North America and Europe. To date there has been little success in developing devices (fish ladders) that enable fish to progress over high dams to upstream spawning areas. Irrigation channels may lead fish to doom by leaving them stranded when the channels are drained, and fish may be destroyed by being jammed against intake screens of power developments. Fluctuations in water level may destroy fish nests by making the water over them too shallow or too deep.

Problems of fisheries conservation also arise from fishing. The greatest problem is that of managing each fishery so as to provide sustained yields of desired species. The solutions of these problems lie in research, public education, legislation and law enforcement, and continuing reevaluation of management procedures. Fishing itself may be destructive; the gear may injure the young while capturing harvestable adults. Fishing may exceed the capacity of a species to maintain itself through reproduction and growth. Although the species may not be exterminated by such overfishing, the fishery may become unattractive economically and for recreation. If it "collapses" entirely, this may bring considerable hardship to the fishermen. Similarly, underfishing is wasteful. In small inland waters, underfishing may destroy the quality of a fishery by leading to overpopulation and thus dense stands of undersized fish, which are unattractive to fishermen. Selective fishing for preferred species, often for predatory ones, may cause coarse, unwanted kinds to usurp the fish-producing capacity of a body of water (*see* POPULATION DYNAMICS).

Some problems of fisheries conservation arise naturally. Included are diseases, natural pollution such as "red tide" of protozoa in marine water and algal "bloom" in fresh water, parasites, predators, and the gradual evolution of lakes toward ponds and dry land and of streams toward sluggish, base-level waterways.

Management measures. Judged historically as well as by the geographic extent of application, legal measures to regulate the taking of fish lead all other forms of fishery management. This is particularly true in the commercial fisheries. Commonly, laws control who may fish, as well as where, when, and how they may fish. These laws regulate the kind, size, and amount of fishes that may be taken during a prescribed period. Legal measures such as antipollution legislation also protect the habitat of aquatic organisms. Because of changing conditions, the efficacy of laws should be continually reviewed.

The artificial propagation of fishes and other aquatic organisms, reputedly practiced in China for several centuries B.C., is a means of increasing the quantities of preferred species. In southeast Asia aquiculture combined with rice culture is an important source of fish for human food. Preferred food fishes in Europe (carp, trout) and North America (trout) are also produced for sale at fish farms or hatcheries. In addition, bait fish are propagated (and sold) extensively in North America. Aquatic farm production methods are also applied to oysters (as food or for pearls), frogs, turtles, ornamental fishes (such as goldfish), and water plants (ornamental, as the lotus, or edible, as the water chestnut).

Some artifically propagated fishes are used in the management of sport and commercial fisheries. Many kinds have been established successfully in waters far removed from their native ranges or in newly created water areas. Still others have been stocked successfully to bolster numbers available to sport fishermen. However, countless others have been placed in waters not suited to them and they soon disappeared, or they have been planted where adequate spawning stocks of the species were already present, in which cases they were wasted or they aggravated an already bad situation of overpopulation. In other situations, the artificially stocked fish have increased so rapidly that they have destroyed or seriously menaced desirable species already present.

Aquatic life in continental lakes, ponds, and streams requires a stable supply of water, with chemical and physical factors varying according to species requirements. Consequently, much effort is [illegible]

vide the prescribed conditions. Often the environment can be improved. Starting in the headwaters of a stream system, land use practices may be adjusted with the water world in mind, for example, by retarding surface runoff and encouraging percolation of water into the ground (*see* LAND USE PLANNING). In the water area itself, chemical, physical, and biological changes can be made to enhance stability, productivity, and yield to fishermen. Pollution and erosion may be controlled and improvement made in food, feeding, and shelter conditions. Species composition, competition, predation, and fishing pressure can be regulated to some extent.

Ponds have been built for many centuries to provide fish and fishing in areas having few natural surface water bodies. In the middle years of the twentieth century, there was a surge of artificial pond or lake construction in the United States and hundreds of thousands of small impoundments resulted from the "farm pond" program, especially in states not bordering the Great Lakes.

Fishery management measures also include educational programs on fish conservation. They likewise encompass basic and developmental research in government and university laboratories. Moreover, these measures include the training of professional fishery scientists to develop and apply the most effective fishery methods.

In general, governments regulate inland, coastal, and boundary-water fisheries. Fishermen are commonly required to buy licenses to fish. In the United States, regulation is primarily at the state level, with the state retaining ownership of the fishes in public waters. In many international waters, international agreements and treaties are used. Often these are in the form of international commissions such as the North Pacific Fishery Commission, Northwest Atlantic Fishery Commission, and the International Great Lakes Fishery Commission.

There are numerous governmental aids to the conservation of fisheries at the local, national, and international levels. Educational programs constitute one of these. Another aid is in the form of exploration for new stocks and the development of more effective gear and methods for handling, storing, manufacturing, shipping, and marketing the

products. There is also direct Federal aid to states for research and the development of fishing. This is particularly true in the United States where certain taxes are earmarked for the purpose (see WILDLIFE CONSERVATION). The Food and Agriculture Organization of the United Nations, as well as certain individual nations, extends international aid and mutual assistance in fishery conservation. [K.F.L.]

Fission, nuclear

The division of an atomic nucleus into parts of comparable mass. The division may take place naturally (spontaneous fission) or under bombardment with neutrons, γ-rays, or other carriers of energy (induced fission). Among light nuclei, the reaction $Be^8 \rightarrow 2He^4 + 0.1$ Mev (5×10^{-14} sec) is the only instance of spontaneous fission, and the reactions $n + Li^6 \rightarrow He^4 + H^3$ and $n + B^{10} \rightarrow Li^7 + He^4$ the only examples of slow-neutron-induced fission. The light and medium-heavy nuclei of other elements are energetically stable against fission.

The energy release on division into two equal parts first becomes systematically positive in the neighborhood of ${}_{40}Zr^{90}$ and rises to about 180 Mev or 3×10^{-4} erg for ${}_{92}U^{235}$. As a result the word fission is generally restricted to these heavier nuclei, all of which are energetically unstable against fission. However, the magnitude of the energy release in fission compared to that in other nuclear reactions (1–20 Mev) is not in itself sufficient to make spontaneous fission occur with any observable probability, as evidenced, for example, by the stability of U^{235}. See NUCLEAR REACTION.

Liquid drop model. The stability of U^{235} is most readily interpreted when the nucleus is compared with a droplet of an incompressible and uniformly charged liquid endowed with a surface tension. Such a droplet is stable against small deformations when the fissility parameter x is less than 1. The

$$x = \frac{(\text{charge})^2}{10(\text{volume})(\text{surface tension})}$$

energy rises as the droplet is deformed until a certain critical deformation is reached. There the droplet has another configuration of equilibrium (Fig 1), but this equilibrium is unstable. In this critical form, the surface tension around the neck of the figure just counterbalances the electrostatic repulsion between the two halves of the droplet, as in the case of stable equilibrium. However, if a slight disturbance moves the two halves slightly apart, the thickness of the neck decreases. Then the attractive force of surface tension loses out to the disruptive tendency of the electrical forces. The two halves of the droplet are accelerated away from each other with ever-increasing velocity, even after the fluid has torn apart (scissioned) into two drops plus tiny residual droplets.

The height of E_f, the energy barrier against fission, has been estimated to lie between 5 and 5.5

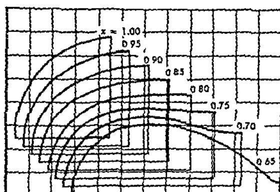


Fig 1. Critical form of unstable equilibrium for uniformly charged liquid drop with fissility parameter x, based on calculations by N. Metropolis and S. Frankel. Only one quarter of each equilibrium shape is portrayed; the rest of each form is obtainable by reflection. Parameter x given approximately in terms of charge number Z and mass number A by the relation $Z^2/47A$.

Mev for ${}_{90}Th^{232}$, ${}_{92}U^{233}$, ${}_{92}U^{235}$, ${}_{92}U^{238}$, and ${}_{94}Pu^{239}$ by irradiating these nuclei with γ-rays of controlled energy and observing that energy at which fission first occurs with appreciable probability. Neutrons are more frequently used as the bombarding agency. In this case the minimum kinetic energy at which appreciable fission occurs is observed and to this fission threshold, E_n, is added the heat of condensation or binding energy, B_n, of the neutron to find the height of the barrier, E_f (Table 1).

The barrier so found refers not to the target nucleus but to the compound nucleus—in this case a nucleus of the same charge but one higher mass number, as for example $U^{235} + n \rightarrow U^{236}$. See NUCLEAR STRUCTURE.

The liquid drop model predicts zero fission barrier for a droplet with the extreme fissility parameter $x = 1$. For droplets of decreasing x values the critical deformation energies steadily increase. For actual nuclei the observed fission barriers show a general upward trend with decreasing values of

Table 1. Fission thresholds and fission barriers found from neutron bombardment*

Target nucleus, T	Mass of compound nucleus, C	Threshold kinetic energy, E_n	B_n	E_f
${}_{90}Th^{232}$	233	1.05	5.06	6.2
${}_{91}Pa^{231}$	232	0.45	5.54	6.0
${}_{92}U^{233}$	234	Slow	6.77	<6.77
${}_{92}U^{234}$	235	0.28	5.25	5.6
${}_{92}U^{235}$	236	Slow	6.44	<6.44
${}_{92}U^{238}$	239	0.92	4.78	5.7
${}_{93}Np^{237}$	238	0.25	5.37	5.6
${}_{94}Pu^{239}$	240	Slow	6.43	<6.43

* All energies are given in Mev and have uncertainties of the order ±0.2 Mev, "slow" refers to neutrons of negligible kinetic energy—thermal energy (0.025 ev) or less; $E_f = B_n + E_n$, critical amount of energy which must be given to this compound nucleus in order to surmount the fission barrier

the nuclear parameter (charge)^2/(mass number) = Z^2/A, although individual nuclei sometimes depart as much as 1 Mev from the general trend. A combination of experimental and theoretical evidence points to barriers for odd-A nuclei on the average roughly 0.5 Mev higher than for neighboring nuclei with even Z and even A. Therefore, insofar as other factors are disregarded, less energy should be required to split the compound nucleus $U^{236} = U^{235} + n$ than $U^{239} = U^{238} + n$. In addition, the heat of condensation set free by the addition of a neutron is greater by about 1 Mev when this neutron completes a pair, as in U^{236}, than when it does not, as in U^{239}. These odd-even effects dominate over any differences between the two uranium isotopes that are systematic in mass number A. Thus to surmount the fission barrier of U^{239} requires about 0.9 Mev more than is set free by addition of a slow neutron, whereas passage over the fission barrier of U^{236} does not require as much as the energy of excitation made available by combination of U^{235} with a slow neutron. Therefore the rare ingredient (1 part in 139) in natural uranium, U^{235}, readily undergoes fission on slow-neutron bombardment. U^{233} and Pu^{239} share this property for similar reasons. One or another of these three substances is required for the functioning of every nuclear chain reaction, whether in a power reactor or in an atomic bomb (*see* ATOMIC BOMB; CHAIN REACTION, NUCLEAR; REACTOR PHYSICS). Important nuclear properties of these principal slow-neutron-fissile substances are summarized in Table 2.

Chain reaction. For a simplified account of fission in its bearing on nuclear chain reactions, it is sufficient to note two facts. First, the act of fission typically releases two or three neutrons which can be used to induce further fissions. For example, when 1000 U^{235} nuclei undergo fission by slow neutrons, they give off varied numbers of neutrons, dividing on the average approximately as follows: 20 (no neutrons); 160 (1 neutron); 340 (2); 310 (3); 130 (4); 40 (5); 0 (6 neutrons). Second, the probability for one of these neutrons to interact with a fissile nucleus in an extended homogeneous medium and produce fission is given by the ratio

$$n_f\sigma_f/(n_1\sigma_1 + n_2\sigma_2 + \cdots + n_f\sigma_f + n_f\sigma_r + n_f\sigma_i)$$

Here n_f is the number of fissile nuclei per cm^3 of the medium and n_1, n_2, n_3, etc. are the numbers of competitive nuclei per cm^3 due to materials such as structural aluminum and graphite for slowing down neutrons.

The cross section σ measures the effective target area (in cm^2 or in 10^{-24} cm^2) of the respective nucleus for interception of a neutron and for bringing about the specified reaction [σ_1, capture by aluminum nucleus; σ_2, scattering by carbon with a loss of energy by the neutron; . . . ; σ_f, fission of U^{235}, for example; σ_r, cross section for the radiative capture reaction $U^{235} + n \rightarrow U^{236}$ (excited) $\rightarrow U^{236}$ (unexcited) + γ-rays]. The cross section depends upon energy. Neutrons of thermal energy (0.025 ev) are most important in the functioning of most chain reactors. At thermal energy the cross section is well known for most materials important for the construction of chain reactors. Therefore the data in Table 2 for fission by thermal neutrons are vital for any first estimate of the neutron economy of a chain reactor. *See* NEUTRON CROSS SECTION.

Fission probability. Appropriate now is a fuller review of the dependence of the probability for fission upon energy, without any attempt to trace out the important consequences of this information for reactor design. The fission probability is best conceived of as a property of a specific energy level of the nucleus. The lowest of these levels have spacings of some tens or hundreds of kev. At higher energies the levels come closer and closer together and have an average spacing of the order of 1 ev in U^{236} at the excitations near 6.4 Mev which are made available by uptake of a slow neutron in U^{235}. Levels at still higher energies are so closely packed that it no longer makes sense to speak of individual levels. Each level, like an independent species of radioactive nucleus, is characterized by a radioactive decay constant A. This quantity gives the probability per second that the

Table 2. Cross sections for neutrons of thermal energy to produce fission or undergo capture in the principal nuclear species, and neutron yields from these nuclei*

Nucleus	Cross section for fission, σ_f, 10^{-24} cm^2	σ_f plus cross section for radiative capture, σ_r	Ratio, $1+\alpha$	No. neutrons released per fission, ν	No. neutrons released per slow neutron captured, $\eta = \nu/(1+\alpha)$
U^{233}	530 ± 8	591 ± 8	1.115 ± 0.020	2.55 ± 0.06	2.29 ± 0.03
U^{235}	576 ± 15	680 ± 10	1.180 ± 0.020	2.47 ± 0.05	2.09 ± 0.02
Pu^{239}	770 ± 15	1108 ± 15	1.439 ± 0.030	2.91 ± 0.06	2.02 ± 0.03
U^{238}	0	2.8		2.4 ± 0.2	0
Natural U	4.1	7.7	1.85	2.47 ± 0.05	1.33

* Data from J. A. Harvey and J. E. Saunders, in *Physics and Mathematics*, ser. 1, vol. 1, Progress in Nuclear Energy, Pergamon Press, 1956. All cross sections in units of 10^{-24} cm^2. Incident neutrons assumed to have maxwellian distribution of energies at 20.4°C.

nucleus will spontaneously undergo transformation. This probability is the sum of partial decay rates A_r, A_f, A_n, A_a, and so on. These magnitudes represent the probabilities per second respectively for dropping to a lower energy level by emission of radiation, for undergoing fission, for sending out a neutron, for emission of an alpha particle, and so forth. The mean life of the energy level is $1/A$ and its half life is $T_{1/2} = 0.693/A$ (*see* HALF-LIFE; RADIOACTIVITY). As a result of the finite lifetime, the energy level is not sharp. For example, neutrons which are emitted by nuclei excited to a particular energy level come off with a distribution in energy. At half maximum intensity, the full width of this distribution is $\Gamma = \hbar A$. Here $\hbar$, the quantum of action, is Planck's constant divided by 2π and is numerically equal to 1.054×10^{-27} erg-sec $= 0.66 \times 10^{-15}$ ev-sec. The total level width is the sum $\Gamma = \Gamma_r + \Gamma_f + \Gamma_n + \Gamma_a + \cdots$ of partial level widths $\Gamma_r = \hbar A_r$, $\Gamma_f = \hbar A_f$, $\cdots$, due to the various competing processes for spontaneous transformation of the given state of the compound nucleus.

The level widths and decay rates belong to the levels themselves. They are unrelated to the formation of the compound nucleus, whether this formation is by the reaction $U^{235} + n$ or otherwise. These widths govern not only the likelihood for a given state of the compound nucleus to break up in any specified manner but also the probability that it will be formed in the first place.

When a neutron has a high energy, say 1 Mev or more, the cross section for fission or for capture of the neutron is approximately expressible as the product of the geometrical area factor π (nuclear radius)$^2 = \pi R^2 \sim 2.4 \times 10^{-24}$ cm^2 for formation of the compound nucleus and the branching factor for the chosen mode of transformation; for example, for neutron-induced fission, the cross section for a neutron to react with a nucleus and produce fission is abbreviated by the symbol $\sigma(n,f)$ and has the approximate value

$$\sigma(n,f) \sim \pi R^2 \frac{\Gamma_f}{\Gamma_r + \Gamma_f + \Gamma_n + \Gamma_a + \cdots}$$

Knowing the magnitude of the partial level widths Γ for the different competing modes of transformation and how they depend upon energy, one can understand from this formula how the cross section for neutron-induced fission depends upon energy (Fig. 2) for neutron energies equal to a substantial fraction of 1 Mev or more.

When the incoming neutron has a low energy, the likelihood of reaction is substantial only when the energy of the neutron is such as to form the compound nucleus in one or another of its resonance levels. The requisite sharpness in "tuning" of the energy is specified by the total level width Γ. The cross section at the center of the resonance is determined by the geometrical size of the nucleus no more than the effective target area of a tuned radio antenna is set by the diameter of the

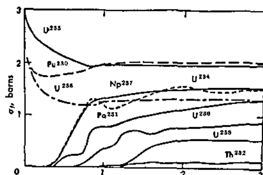


Fig. 2. Cross section for neutron-induced fission (1 barn = 10^{-24} cm^2) as a function of neutron energy. (*Based on D. J. Hughes and R. B. Schwartz, Neutron Cross Sections, Brookhaven Natl Lab. AEC-BNL 325, Suppl. 1, 1957*)

wire; it is governed instead in both cases by the square of the wavelength of the incoming radiation. For neutrons of energy E close to an isolated resonance of spin 0 and of resonance energy E_{res}, the cross section is given by the formula of G. Breit and E. P. Wigner

$$\sigma(n,f) = (1 \text{ ev}/E)650{,}000 \times 10^{-24} \text{cm}^2 \frac{\Gamma_n \Gamma_f}{(E - E_{res})^2 + (\frac{1}{2}\Gamma)^2}$$

The high cross section of U^{235} to take up a slow neutron and to undergo fission makes it possible for U^{235} to compete successfully for slow neutrons against 139 times as many U^{238} nuclei in a reactor fueled by natural uranium (Table 2).

Fission fragments. A number of processes accompany fission. Let U^{235} be exposed to a sufficiently intense irradiation by neutrons of thermal energy to produce 10^6 cases of fission. Then there occur at the same time 184,000 cases of simple uptake of a neutron to release 6.4 Mev of γ-ray radiation and form U^{236} in its ground state. For the 10^6 nuclei that divide, the fission fragments come off with about 168 Mev of energy. They have also an internal excitation which they give off in much less than 10^{-5} sec in the form of about 6 γ-rays, mostly in the interval 0.5–2.5 Mev, and of total energy about 5 Mev.

The neutron-to-proton ratio of the fragments made out of $n + U^{235} \rightarrow U^{236} \rightarrow 2$ fragments plus 2 or 3 neutrons is unbalanced ($141.5n/92p$) compared to the neutron-to-proton ratio of normal stable nuclei in the region of masses from $A = 90$ to $A = 150$. As a result each fragment undergoes on the average about 3 β-decays before it settles down to a stable nucleus. In the 6 β-decays per fission the β-particles (electrons) take away about 6 Mev per fission, the accompanying 6 antineutrinos take ~10 Mev, and associated γ-rays take about 5 Mev. About 2000 of the 10^6 acts of fission are accompanied by the emission of an α-particle between 5 Mev and 25 Mev, and of average about 20 Mev, emitted preferentially at

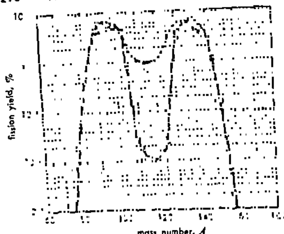


Fig 3. Mass distribution of fission fragments formed by neutron-induced fission of $U^{235} + n = U^{236}$ when neutrons have thermal energy, smooth curve (*Plutonium Project Report, Revs. Mod Phys, 18* 539, *1946*) and 14-Mev energy, dashed curve (*based on R W. Spence, Brookhaven Natl Lab, AEC-BNL(C-9), 1949*). Quantity plotted is 100 × (number of fission decay chains formed with given mass)/(number of fissions).

gles to the line of separation of the fission fragments.

Of the 2.5 neutrons given off per act of fission, the average neutron has a kinetic energy of about 1 Mev with respect to its parent fragment and another 1 Mev by reason of the motion of this fragment, or a total of about 2 Mev. The two velocities add vectorially, and each has considerable spread so that the energy distribution of these neutrons is very broad. These neutrons are given off in a time that is much less than 10^{-7} sec and are therefore called prompt neutrons.

Additional neutrons are given off with an appreciable time delay through another mechanism by fission fragments which have been brought to rest. These delayed neutrons amount in the case of the reaction U^{235} + thermal neutron to about 0.7% of all the neutrons given off in fission. Though small in number, they are quite important in stabilizing nuclear chain reactors against sudden minor fluctuations in reactivity. The mechanism of delayed neutron emission is a simple deviant from the normal process of β-decay of a fission fragment ${}_ZA \rightarrow$ ground state or a moderately excited state of ${}_{(Z+1)}A \rightarrow$ normal decay chain history. Occasionally the β-decay leads to a state of the daughter nucleus ${}_{(Z+1)}A$ of excitation great enough to surpass the neutron binding energy. In that case a neutron "boils off" in a time less than 10^{-14} sec and there is left the nucleus ${}_{(Z+1)}(A-1)$, almost always in its ground state. The time delay is associated with the β-decay of the mother nucleus. Examples include Br^{87}(54.5 sec) $\rightarrow Kr^{87}$ (highly excited) $\rightarrow n$(average energy 0.25 Mev) + Kr^{86}; I^{137} (21.8 sec) $\rightarrow Xe^{137}$ (highly excited) $\rightarrow n$(average energy 0.56 Mev) + Xe^{136}; and other delayed neutron periods of 6.0 sec, 2.2 sec, 0.50 sec, and 0.18 sec. *See* NEUTRON; NEUTRON, DELAYED; THERMAL NEUTRON.

Division into two fragments of equal mass is about 600 times less probable than division into the most probable choice of fragments when U^{235} is irradiated with thermal neutrons. When the energy of the neutrons is increased, symmetric fission (Fig. 3) becomes more probable.

Fission has been observed to be produced not only by neutron and γ-ray irradiation, but also by protons, deuterons, α-particles, and mesons. When the bombarding energy is very great, numerous neutrons are often given off (spallation) before fission itself ensues. *See* SPALLATION REACTION.

Spontaneous fission sets a fundamental limit to the stability of very heavy nuclei. Nuclei of odd mass number decay more rapidly by spontaneous fission than neighboring even-even nuclei by a factor of 10^2–10^4. The most stable nucleus of charge 92 (U^{238}) has a half-life against spontaneous fission of about 10^{16} years. This life drops by a factor of about 10^4 for each increase of two units in the nuclear charge, becoming about 1 year for $Z = 100$ (Fm^{254}). [J.A.W]

Bibliography: E. U. Condon and H. Odishaw (eds.), *Handbook of Physics*, 1958.

FitzGerald-Lorentz contraction

The contraction of a moving body in the direction of its motion. In 1892 G. F. FitzGerald and H. A. Lorentz proposed independently that the failure of the Michelson-Morley experiment to detect an absolute motion of the earth in space arose from a physical contraction of the interferometer in the direction of the earth's motion. According to this hypothesis, as formulated more exactly by Albert Einstein in the special theory of relativity, a body in motion with speed v is contracted by a factor $\sqrt{1 - v^2/c^2}$ in the direction of motion, where c is the speed of light. For a more detailed discussion, *see* RELATIVITY. For a discussion of the Michelson-Morley experiment *see* LIGHT. [E.L.H.]

Flabellifera

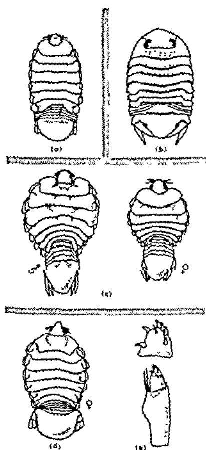
The largest and morphologically most generalized suborder of isopod crustaceans. The biramous uropods are attached to the sides of the abdomen and may with the last abdominal segment form a caudal fan. More than 1400 species and over 10 families are known.

Members of the family Cirolanidae, such as *Cirolana*, *Exocirolana*, and *Eurydice*, are actively swimming predators and scavengers with biting mouthparts. They feed on dead and living animals and are notorious for their attacks on the bait of line fishermen and on fishes caught in nets. The legs of bathers are also sometimes bitten. The largest known isopods, species of *Bathynomus*, are inhabitants of the deep sea and may reach a length of more than 350 mm. The endopods of the pleopods bear tufts of respiratory filaments which aid in respiration. In other Cirolanidae, *Anuropus* and *Branchuropus*, the

uropods do not form a tail fan, but are similar in form and position to the pleopods.

Members of the families Corallanidae and Excorallanidae are sometimes found parasitizing fishes, but are often free-living. The mandibles and first maxillae form hooklike piercing organs in the Excorallanidae, while the first maxillae may have a somewhat similar form in the Corallanidae. The maxillipeds in the above families are unmodified, but in the fish parasites of the families Aegidae and Cymothoidae they are reduced, and end in hooks. The Aegidae feed by sucking blood, and after feeding, the thoracic somites become greatly distended.

The Cymothoidae are more highly adapted for a parasitic life than other Flabellifera. All the thoracic legs have large uncinate dactyls for holding on to the host. The cymothoids, *Anilocra* and *Nerocila*, are parasitic on the body surface. *Lironeca*, *Cymothoa*, *Codonophilus*, *Irona*, and others are found in the mouth and gill cavities, while *Orozeuktes* and *Artystone* may even bore through the skin and enter the body cavity. Most, if not all, of the cymothoids are protandrous hermaphrodites.



Examples of Flabellifera, diagrammatic. (a) *Lironeca epimerias* Richardson. (b) *Tecticeps renoculis* Richardson. (c) *Ichthyoxenus jellinghausii* Herklots, male and female (from Proc. U.S. Natl. Museum). (d) *Braga patagonica* Schioedte and Meinert, female. (e) *Rocinela simplex* Chilton.

When the young leave the mother's marsupium, they swim freely and eventually settle on a fish. The eyes are then reduced, many of the natatory setae are lost, and male sex organs develop. Later, the testes undergo regression and ovaries ripen. The mature female retains the appendix musculinum of the second pleopod as evidence of its former masculine function.

The gribbles, Limnoriidae, are well known for the damage they cause to marine structures of wood. In addition to the wood-boring species of *Limnoria* and *Paralimnoria*, a number of species of *Limnoria* (subgenus *Phycolimnoria*) burrow into the holdfasts of kelps.

In the family Sphaeromidae, the inner branch of the uropod is immovable. Sphaeromids can curl up into a pill-like form, or fold like a closed book. In some sphaeromids, the eggs are incubated in pouches formed by paired invaginations of the ventral body wall. Some species of *Sphaeroma* are economically important in tropical seas because of their habits of boring into wood, clay, and rock.

Isopods of the family Serolidae are greatly flattened forms which live partially buried on sandy bottoms. See ISOPODA. [T.E.B.]

Bibliography: H. J. Hansen, Cirolanidae et familiae nonnullae propinque Musei Hauniensis, *Kgl. Danske Videnskab Selskabs Skrifter, Naturvidenskab, Math. Afdel*, V, 3, 1890; A. Milne-Edwards and E. L. Bouvier, *Les Bathynomes*, Mem. Museum Comp. Zool. Harvard College, 27 (2), 1902; T. Monod, Sur quelques crustacés aquatiques d'Afrique (Cameroun et Congo), *Rev. zool. botan. africaines*, 21(1):1–36, 1931.

Flame

A reaction front (or wave) in a gaseous medium into which the reactants flow and out of which the products flow. Solid or liquid particles may also be carried in by the gas and provide fuel for the flame; and sometimes the product gases contain solid particles, such as soot.

Flames produce heat, and usually light. Hydrogen flames, for example, are almost invisible, although they do radiate in the infrared and ultraviolet regions of the spectrum. Nowadays, flames are used mostly for heat and very little for light. The greater part of the heat can be extracted only by direct conduction from the hot gases, because the radiant effect is small. For this reason the vigorous flames in a fireplace give very little warmth unless the air of the room is circulated through a duct in the firebox.

A seeming paradox is worthy of note, however. A faint, luminous phenomenon, known as a cool flame, is observed when, for example, a mixture of ether vapor and oxygen is slowly heated. Such flames are not really cool; they do produce some heat. But, unlike ordinary flames that advance from layer to layer in the gas mainly by heat conduction, cool flames proceed by diffusion of reactive molecules, or free radicals, which initiate chemical processes as they go. The light from cool flames is

a chemiluminescence that originates from excited molecules created by the reaction. *See* CHEMILUMINESCENCE.

Combustion flames, in which oxygen or some other oxidizing substance combines with a fuel, are the more common. But decomposition flames, in which a molecule such as ozone, O_3, breaks down into simpler molecules, in this case oxygen, O_2, are also known. Not all combustion or decomposition reactions, however, produce a flame, even though they may evolve heat. For example, charcoal undergoes surface combustion; and, although it glows, there is no flame. *See* COMBUSTION.

Premixed and diffusion flames. Gaseous fuels can be premixed with air or oxygen, in which case the mixture can be fed to a flame holder and burned in a very efficient manner. In a gas stove or a laboratory burner, an injection nozzle for fuel and ports for air are arranged so that air is drawn in by aspiration and mixed with fuel before it emerges at the flame holder.

Premixed flames can be either laminar or turbulent. The laminar flame front is smooth and sharply defined, and both the approach and exhaust flows are streamlined. The turbulent flame, on the other hand, may be described as bushy and is often quite noisy The flame front fluctuates rapidly and is filled with swirls and eddies. Such turbulence results partly from the flow in the burner tube and partly from the flame itself. Turbulent flames are associated with high-speed flow, and, because of the large convoluted burning area, they are very efficient in terms of heat production per unit volume. Turbulent burning is therefore desirable in many practical applications.

Liquid and solid fuels burn in a so-called diffusion flame. A candle provides a good example. The fuel—wax, in this case—is melted and vaporized by the heat of the flame and emerges as a steady

Because certain parts are fuel-rich, hydrocarbon

On the other hand, some diffusion flames do not form carbon; the methyl alcohol flame, for example. Oil burners, coal furnaces, liquid-fueled rockets—in fact, the great majority of technical combustion devices—involve diffusion flames.

Bunsen flame. The bunsen burner has been extensively studied because it illustrates important concepts about flames. With a *fuel of natural* gas (methane, CH_4), or certain other hydrocarbons such as ethylene, C_2H_4, and with the air ports completely closed, a bright yellow flame is produced. This is the diffusion flame, *in which the air is* supplied entirely as secondary air that diffuses from the atmosphere. As the air ports are gradually opened, the yellow color diminishes and finally disappears. The flame then consists of three distinct regions.

The innermost region is dark. It is simply the cold, unburned gas. If a match head is suspended



Fig. 1. Experiment showing that the flame cone is hollow. The match head will not light.

The dark region of unburned gas is capped by a bright bluish-green, cone-shaped mantle, less than one millimeter thick. In this narrow layer, the fuel reacts with the premixed air. In the usual bunsen flame, however, the amount of premixed air is insufficient for complete combustion. Therefore, the gas emerging from the reaction zone still has fuel value. Analysis shows it to contain carbon monoxide, CO, and hydrogen, H_2, but none of the original methane. This gas can react with more oxygen to give carbon dioxide, CO_2, and water, H_2O.

The additional oxygen needed for combustion of gases emerging from the inner cone is supplied by secondary air from the atmosphere. Above the inner cone, therefore, is a pale bluish-violet diffusion flame. This is called the outer cone.

By means of two concentric tubes, called a Smithell's burner, the inner and outer cones can be separated. The gas between the cones *in this* burner emits no light. The light from a methane flame is emitted only by regions in which chemical reaction is taking place; none of the light comes from the hot combustion products. Spectroscopy shows that the flame radiation is produced by short-lived radi*cals* such as C_2 and CH.

When the ports of the burner are opened to admit more primary air, the outer cone wanes and the inner cone becomes darker. Eventually, when enough primary air is available for *complete* combustion, the outer cone disappears.

Flameholder. A flame is stabilized on a burner tube because of heat flow to the cold rim. Drainage of heat lowers the flame speed at this point and, as a result, the front becomes anchored near the rim. If the front should momentarily move further from the rim, less heat would drain, and the front would speed up and return toward its original position near the rim. On the other hand, if the front should

(a)

(b)

(c)

(a) Premixed ethylene-air flame, fully aerated
(b) Diffusion flame of ethylene in air.
(c) Premixed ethylene-air flame, partially aerated
(d) Ethylene-air flame with nitric oxide added.
(e) Flame of acetylene and nitrous oxide at a pressure of about 3 mm Hg.
(From A. G. Gaydon, Flames: their structure and radiation, Endeavour, 10(37):17–21, 1951)

(d)

(e)

shift toward the rim, it would lose more heat, the *burning velocity would drop, and the gas stream* would drive the flame back to its original place. Any solid surface may act in this way as a flame holder.

The height of the flame cone above the burner rim adjusts itself within wide limits to accommodate the gas flow. The surface of the cone is actually curved as a result of the parabolic flow distribution in the burner tube. With larger flow rates, the cone height increases, and with smaller rates, it decreases. However, above a certain limiting flow, the flame lifts off the burner and blows out; this is called blow-off. On the other hand, if the flow is sufficiently reduced, the cone becomes almost flat. A further decrease at this point will cause the flame to strike back into the burner tube; this is called flashback.

Between the flashback and blow-off limits, the cone assumes a height such that the volume rate of flow divided by the surface area of the cone retains a constant value, which is called the burning velocity and denoted by S_u.

Burning velocity. This is a basic property of a combustible gas mixture. It is affected by external factors only when heat is drained directly from the reaction zone near a flame holder. If, by special arrangements, a gas stream of uniform velocity over the cross section is produced, the flame front can be made almost perfectly flat. Under such a condition, the flame is stable only when the linear flow rate is exactly matched to the value of S_u. Burning velocity always increases with rising gas temperature and usually shows a dependence on initial pressure also; but the behavior in this respect differs for various mixtures.

Burning velocities range through wide limits; methane-air, for example, has a velocity of only 35 cm/sec, whereas certain hydrogen-oxygen mixtures burn at a rate of 1200 cm/sec. No simple expression relates burning velocity and other properties of the mixture.

The theory of flames is, in fact, very complex. The properties influencing S_u are the heat conductivity, the diffusivity, the flame temperature and the reaction rate. Because diffusivities of the various molecular species involved in the reaction differ and *because the concentration of these species* both affects, and is affected by, the rate of reaction, the mathematical analysis of what takes place in the burning zone is a formidable problem. The principles are well understood, but the interdependent factors are so complex that no general solution of the differential equations is possible. They have been solved in certain cases by the use of computing machines.

In spite of actual complications, there is no mystery in the mechanism of flame propagation in general. Both heat conduction and diffusion of reactive species serve to start reaction in successive layers of gas. When very mobile and reactive H atoms are involved, their diffusion from the burned to the unburned region dominates the process, but

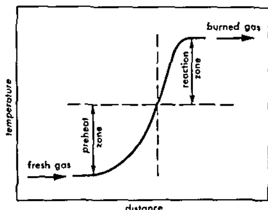


Fig. 2. Temperature distribution through the flame front.

usually both diffusion and heat conduction are important. By considering only the latter and disregarding the effect of diffusion (which will never be strictly permissible) a good intuitive picture of the flame can be formed. Conduction of heat depends on the temperature gradient; and, by quite straightforward reasoning, it can be shown that the temperature distribution through the flame will always be S-shaped. This conclusion is indeed confirmed by actual thermocouple measurements (Fig. 2). The net rate at which heat is received by an element of gas depends on the curvature of this temperature profile. Where it is curving upward, at the lower part of the S, the gas is receiving more heat than it is losing; and where it is curving downward, in the upper portion of the curve, the gas is losing more heat than it is receiving. The first region may be considered as the preheat zone. Here, the gas is still not hot enough to react at an appreciable rate. In fact, most of the reaction occurs only near *the region of highest temperature* because the reaction rate, following the exponential Arrhenius law, increases very sharply as the temperature rises. Opposing the effect of temperature as the reaction nears completion, however, is the slowing effect resulting from the depletion of reactant concentrations. According to a much simplified picture, the flame propagates by the flow of *heat from the reaction zone to the preheat zone.* An element of gas passing through the flame is heated up in the preheat zone to a temperature where it begins to react. Then, as it passes on through the reaction zone, it self-heats and, at the same time, hands back the extra heat it had previously received to the succeeding element of gas, now in the preheat zone. Thus the flame is maintained in a steady state. *See* EXPLOSION AND EXPLOSIVE; FUEL; KINETICS (CHEMICAL). [W.E.G.]

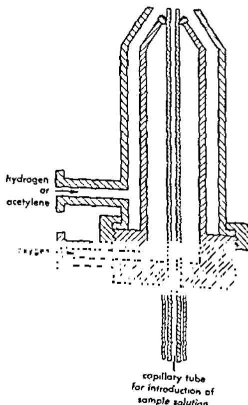
Bibliography: A. G. Gaydon and H. G. Wolfhard, *Flames. Their Structure, Radiation and Temperature,* 1953; W. Jost, *Explosion and Combustion Processes in Gases,* 1946; B. Lewis and G. von Elbe, *Combustion, Flames and Explosions of Gases,* 1951.

Flame photometry

A branch of spectrochemical analysis in which samples in solution are excited to luminescence by introduction into a flame. Flame photometry is particularly useful in determining small amounts of lithium, sodium, and potassium in solution. Only small samples are required, and an analysis usually can be carried out in relatively short time.

In the Lundegardh vaporizer, a compressed-air aspirator vaporizes the solution within a chamber. The smaller droplets are carried along into the fuel-gas stream and then to the burner orifice. As each droplet passes through the flame, the solvent is evaporated, and the solutes are vaporized, dissociated, and optically excited. This system usually is used with slow-burning fuel mixtures, such as illuminating gas and air. The burner may be similar to a meker burner. Most flame photometry in the United States, however, now is done with the Beckman burner, which aspirates solution directly into a hydrogen-oxygen or acetylene-oxygen flame.

Since flame temperatures are comparatively low, only the more volatile compounds are vaporized from the residue. Also, only the less stable molecules are dissociated, and only the more easily excited spectral lines and bands are emitted. Flame line-emission spectra are thus relatively simple, although molecular band spectra caused by the fuel or metallic oxides sometimes cause spectral interference.



Beckman aspirator-burner for use in flame photometry.

The line or band of interest is isolated with a monochromator, and its intensity measured photoelectrically. The size and type of monochromator required depend on the spectral region studied and the complexity of the spectrum. For the 4000–7700 A region, a small or medium prism instrument often is adequate, although a grating monochromator would be preferable. In the 7700–8600 A range, which is used in rubidium and cesium analyses, the grating gives much superior performance.

Flame photometry is more accurate, precise, and sensitive than wet chemical methods for the determination of small amounts of alkali metals, and it rivals wet methods in the determination of some other elements. Standard deviations are usually 1–5% of the amount present. Sensitivity limits for various elements go as low as fractions of a microgram per liter to several hundred milligrams per liter, depending on the element, the solvent, the solute composition, the portion of flame used, the resolving power of the monochromator, and the sensitivity and signal-to-noise ratio of the detection system. Since flames are steady, internal standards are not needed to compensate for fluctuations. Matrix effects are usually compensated for by comparing the sample with synthetic standards prepared in a matrix simulating that of the sample or by using the standard-addition technique. *See* ATOMIC STRUCTURE AND SPECTRA; SPECTROCHEMICAL ANALYSIS. [C.F.]

Bibliography: F. Burriel-Martí and J. Ramírez-Muñoz, *Flame Photometry*, 1957.

Flameproofing

The process of treating materials so that they will not support combustion. Although cellulosic materials such as paper, fiberboard, textiles, and wood products cannot be treated so that they will not be destroyed by long exposure to fire, they can be treated to retard the spreading of fire and to be self-extinguishing after the igniting condition has been removed.

Treatments. Numerous methods have been proposed for flameproofing cellulosic products. One of the simplest and most commonly used for paper and wood products is impregnation with various soluble salts including ammonium sulfate, ammonium phosphate, ammonium sulfamate, borax, and boric acid. Special formulations are often used to minimize the effects of these treatments on the color, softness, strength, permanence, or other qualities of the paper.

For some applications these treatments are not suitable because the salts remain soluble and leach out easily on exposure to water. A limited degree of resistance to leaching can be achieved by the addition of latex, lacquers, or waterproofing agents. In some cases the flameproofing agent can be given some resistance to leaching by causing it to react with the cellulose fiber (for example, urea and ammonium phosphate).

Leach-resistant flameproofing may also be obtained by incorporating insoluble retardants in the

paper during manufacture, by application of insoluble materials in the form of emulsions, dispersions, or organic solutions, or by precipitation on, or reaction with, the fibers in multiple-bath treatments. The materials involved are of the same general types as those used for flameproofing textiles and include metallic oxides or sulfides, and halogenated organic compounds. Because of the higher material cost and more elaborate equipment required, such treatments are used on paper only when unusual conditions justify their cost.

Action of retardants. Various theories have been proposed to explain the flameproofing action. In general, these are quite similar to those considered in the flameproofing of textiles. Two types of combustion may be involved. In one, the volatile decomposition products burn with a flame; in the other, the solid material undergoes flameless combustion, or afterglow. In general, the alkaline types of retardants are effective in preventing afterflame and the acid types control afterglow. A few, such as ammonium phosphates and halogenated products reduce both flame and glow.

The theory for prevention of afterflame having most support is a chemical one which holds that effective flame retardants direct the decomposition of the cellulose when heated so as to minimize the formation of volatile flammable products and increase the amount of water and solid char formed. The prevention of afterglow is usually attributed to a modification of the flameless combustion to make it less exothermic and therefore incapable of maintaining itself, for example, by formation of carbon monoxide instead of carbon dioxide. Other theories include the formation of a coating or froth which excludes oxygen and smothers the combustion, the formation of nonflammable gases which dilute the flammable products and exclude oxygen, and the thermal theories which hold that heat is dissipated by endothermic reactions or by conduction. It is probable that these actions do contribute in some cases.

The susceptibility of paper or fiberboard to combustion may also be modified by the introduction of noncombustible but otherwise inert materials, such as mineral fillers, asbestos, and glass or ceramic fibers, into the paper or board during manufacture. The high proportions required result in products with properties determined largely by noncombustible components rather than by cellulosic fibers.

The behavior of paper, fiberboard, or wood products exposed to fire will depend also on structural factors. The density of the material and the ratio of mass to surface exposed will affect the ease of ignition and the rate of spreading or penetration by the fire. Fire-resistant coatings will minimize the spread of flame. Heat-reflective coatings or insulating coatings will prevent or retard the material beneath them from reaching the ignition temperature. Intumescent coatings are formulated to form a thick insulating foam when exposed to fire.

A wide variation exists in methods for evaluating the efficiency of a flameproofing treatment or of suitability for a given use. Paper is usually tested by igniting with a flame of specified character for a predetermined period of time and then noting duration of afterflame and of afterglow and the amount of char length or char area. Structural materials are tested by other methods which involve features of construction in addition to those of composition. Methods have been standardized by the Technical Association of the Pulp and Paper Industry (TAPPI), American Society for Testing Materials (ASTM), and various governmental agencies.

For a discussion of textile flameproofing, *see* TEXTILE CHEMISTRY. *See also* COMBUSTION; FIRE EXTINGUISHER; FLAME. [T.A.H.]

Flamingo

Any of 6 species of large, long-legged, long-necked, heronlike birds of the family Phoenicopteridae. They are characterized by a short, strong bill, with a sharp bend downward in the middle. When the bird is feeding, the head is pulled back toward the



The flamingo, Phoenicopterus ruber; length to [illegible] (From P. Martin Duncan, ed., Cassell's Natural [illegible] Cassell & Company, Ltd., London)

body in a shoveling motion, thus [illegible] through the bill in searching for small [illegible] which comprise the flamingo's major [illegible] gos nest in large colonies on mud flats. [illegible] occur in tropical America. The [illegible] can flamingo, *Phoenicopterus* [illegible] times along the Florida coast [illegible] *See* CICONIIFORMES.

Flaser rock

A term applied to sheared [illegible] formed rocks that have not [illegible] stage of granulation [illegible] roughly synonymous [illegible] cataclasite. Flaser gab[illegible] applied to cataclasti[illegible]

parent rock can be recognized in numerous undestroyed lensoid fragments lying between belts and streaks of highly crushed rocks. The alteration often starts by a granulation along the edges of the grains which thus become surrounded by a corona of broken fragments (mortar structure). In more advanced stages the granulated part forms patterns of streaks bending and swirling between the elongated patches of more undestroyed rock. As these relic patches become smaller and the granulated matrix larger, these rocks pass by mylonitic augen gneisses into true mylonites Protoclastic gneisses are petrographically similar but are supposed to have developed through flow of a viscous magma during the final stages of igneous crystallization. *See* METAMORPHIC ROCKS; MYLONITE. [T.F.W.B.]

Flash point

The temperature at which a flash of flame is first observed when a small lighted flame is passed over a sample of a flammable substance in a cup. The flash point is an important test for fire and safety regulations. So also is the fire point ...

the oil and are applied to solvents, distillate fuels, lubricating oils, and many other petroleum products. *See* OIL ANALYSIS. [M.SO.]

Flash welding

A form of resistance welding that is used for mass production. The welding circuit consists of a low-voltage, high-current energy source (usually a welding transformer) and two clamping electrodes, one stationary and one movable.

The two pieces of metal to be welded are clamped tightly in the electrodes, and one is moved toward the other until they meet, making light contact (Fig. 1*a*). Energizing the transformer causes a high-density current to flow through small areas that are in contact with each other (Fig. 1*b*).

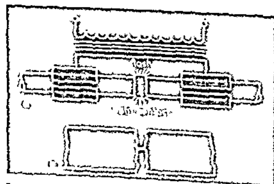


Fig. 1. Elements of flash-welding operation. *(a)* Two pieces of horizontally clamped metal are brought together to make a light contact. At that time the welding circuit is closed to the *transformer*. *(b)* Current starts flowing when small areas of the workpiece make initial contact. (*From A. L. Phillips, ed., Welding Handbook, 4th ed., American Welding Society, 1957*)

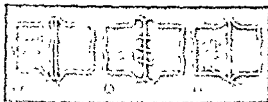
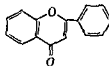


Fig. 2. Flash-weld quality is indicated on the surface *(a)* Satisfactory heat and upset. *(b)* Cracks caused by insufficient heat. *(c)* Insufficient heat or insufficient upset force, or both. (*From A. L. Phillips, ed., Welding Handbook, 4th ed., American Welding Society, 1957*)

Flashing starts and the movable workpiece must accelerate at the proper rate to maintain an increasing flashing action. After a proper heat gradient has been established on the two edges to be welded, an upset force is suddenly applied to complete the weld. This upset force extrudes slag, oxides, molten metal, and some of the softer plastic metal, making a weld in the colder zone of the heated metal (Fig. 2). Flash and upset material are frequently removed, although when the design permits, a portion of the upset material may be retained as a reinforcement. *See* RESISTANCE WELDING. [E.J.L.]

Flavone

One of a number of widely distributed pigments that occur in the plant kingdom. They are chemically related to the parent compound, flavone, which has been isolated from the primrose. The



Flavone

majority of the pigments are yellow and are derivatives of flavone that occur in the plant in combination with sugars or tannic acid.

Of all the natural pigments that can be used as dyestuffs, the flavones are by far the most widely distributed in nature. One of the members of this group, luteolin, the main coloring matter of the herbaceous plant known as weld (*Reseda luteola*), is said to be the oldest European dyestuff known. Some of the flavone dyestuffs that are still significant economically are weld, young and old fustic, and quercitron bark. The use of these, however, is largely confined to backward countries in which they abound.

Certain flavone derivatives exert an effect on the permeability of capillary walls similar to the effect of crude water-soluble extracts of citrus fruit peels containing an active principle designated vitamin P (A. Szent-Gyorgy, 1936). Three which have been widely studied and are marketed as drugs are rutin, quercetin, and hesperidin. These compounds

have been used clinically in the treatment of disease states characterized by capillary bleeding associated with increased capillary fragility. Evidence for their efficacy is not conclusive. *See* DYE.

[S.M.K.]

Bibliography: R. C. Elderfield, *Heterocyclic Compounds*, vol. 2, 1951.

Flax

The flax plant, *Linum usitatissimum*, is the source of two products: flaxseed for linseed oil and fiber for linen products. Plants with two distinct types of growth are used for seed and fiber production (Fig. 1). But fiber is sometimes a by-product of linseed production, and seed is an important by-product of fiber flax. *See* SEED (BOTANY) Flax is an annual plant with a stem about ⅛ in. in diameter. *See* STEM (BOTANY). In a dense stand, it does not branch. The flax flower has five petals which are blue, white, or pale pink. The color of the flower depends on the variety (Fig 2). Seeds of most varieties are light brown, but yellow seeds are common among seed-flax varieties. Usually flax is planted in the early spring and harvested 3–4 months later. In hot climates, seed flax is grown as a winter crop Research has shown that seed flax depletes the soil no more than a crop of wheat.



Fig 2. Top of a seed flax plant showing buds, flowers, and seed capsules. (USDA, ARS)

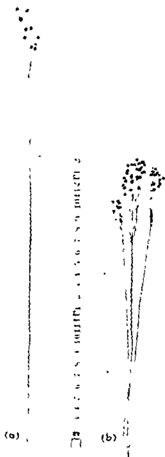


Fig. 1. (a) Fiber flax. (b) Seed flax. (USDA, ARS)

Cultivation and harvesting of fiber flax. The flax plant requires deep, rich, well-plowed soil and a cool, damp climate. Prematurely warm weather affects the growth and the quality of the fiber. Level land with a plentiful supply of water is normally used for fiber flax. Flax is usually grown only once every 7 years on a given piece of land, this practice having been shown to be desirable for high yields Other crops are grown in between in a regular crop-rotation program.

In northern latitudes flaxseed for fiber flax are sown by hand in April or May. When the plants are a few inches high, weeds are pulled by hand with extreme care to avoid injury to the delicate sprouts. In three months, the plants develop straight, slender stalks, 2–4 ft in height, with tapering leaves and small flowers. The variety with blue flowers yields fine fiber. The white-flowered plants produce a coarse but strong fiber. *See* FLOWER (BOTANY).

By the end of August, flax turns a brownish color, which indicates that the plants are almost mature and ready for harvesting. There can be no delay at this stage because the fiber soon loses its prized luster and soft texture. The plants are usually pulled out of the ground by hand; however, this operation can be performed efficiently with ma-

chinery. If the stalks of the plants are cut, the sap is lost and the quality of the fiber is lowered. Therefore, the stalk must be kept intact, and the tapered ends of the fiber must be preserved, so that a smooth yarn may be spun. The stalks are tied in bundles, called beets, in preparation for extraction of the fiber.

Flax for fiber. The best known use for flax fiber is in the manufacture of linen fabrics; other uses are linen thread, linen twine, toweling, and canvas (*see* FIBER, NATURAL; LINEN; TEXTILE). Seed from fiber flax is used for planting or for oil. The major producer of flax fiber is Russia, but the world's best fiber comes from Belgium and adjoining countries. Most Irish linen is manufactured from fiber produced in Belgium. The United States production of flax for fiber has dwindled in recent years from about 18,000 acres in 1942 to none in 1958.

Flax for seed. Less flax is planted for seed than for fiber, and the varieties used produce short straw and a relatively high yield of seed with 32–44% oil (*see* FAT AND OIL, NONEDIBLE). Planting and harvesting are often done by machines and frequently weeds are controlled with herbicides. *See* AGRICULTURAL MACHINERY; HERBICIDE.

In 1958 about 97% of the flaxseed acreage in the United States was in Minnesota, North Dakota, and South Dakota. The world's major producers of flaxseed are the United States, Argentina, Canada, Russia, and India.

The principal uses of linseed oil are in the manufacture of paint, varnish, linoleum, and oilcloth; but it is also used in such products as printer's ink, patent leather, imitation leather, and core oil for making sand forms for metal casting. The linseed cake which remains after the oil is extracted is a high-protein livestock feed. Only a relatively small amount of the straw from seed-flax varieties is used. This is used in cigarette and other fine papers. [E.C.N.; J.O.CU.]

Flax diseases. Of the diseases of flax, the more common are: rust, *Melampsora lini;* wilt, *Fusarium oxysporum lini*, pasmo, *Sphaerella linorum;* browning or stem break, *Polyspora lini*, seedling blight and root rot caused by many species of fungi that belong to the genera *Alternaria, Aphanomyces, Colletotrichum, Fusarium, Phoma, Pythium, Rhizoctonia, Thielavia*, and others; viruses such as aster yellows and curly top; and nonparasitic diseases induced by heat, mineral deficiency, and toxic substances. *See* FUNGI; PLANT, MINERAL NUTRITION OF; PLANT VIRUS.

The importance of these diseases varies greatly with cultural practices, cropping sequence, environmental conditions, quality of seed, and varieties grown. Diseases that are destructive in one region

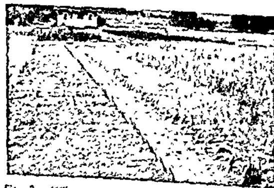


Fig 3. Wilt resistance plot, Minnesota. Each stake marks a line of flax. On left most of plants killed by wilt whereas those on the right are resistant. (Minnesota Agricultural Experiment Station)



Fig. 4. Blight of flaxseed caused by several fungi. (Minnesota Agricultural Experiment Station)

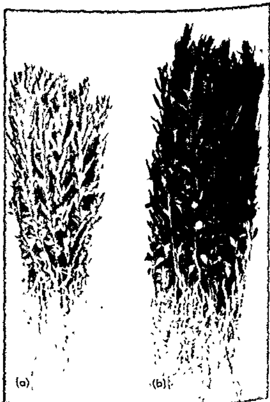


Fig. 5. Effect of seed treatment on the control of seedling blight in flax. (a) Nontreated. (b) Seed treated with an organomercuric compound. (Minnesota Agricultural Experiment Station)

may be of minor importance in another. Thus, pasmo is a major problem in the upper Mississippi Valley, but is of minor importance in California, and it is unknown in Mexico. Likewise, stem break is a troublesome disease in many parts of Europe, but it is of little importance in the United States.

At the beginning of the twentieth century, flax wilt was the limiting factor in flax production in the United States. At this time it was discovered that this disease was caused by a fungus, *Fusarium* sp., and that it was possible to select wilt-resistant lines of flax, if the selections were made on infested, "wilt sick," soil. As a consequence, wilt disease is now controlled by growing resistant varieties (Fig. 3).

Rust is the most conspicuous and potentially the most dangerous disease of flax; it occurs throughout the important flax-growing regions of the world. This fungus comprises numerous distinct parasitic races and no variety of flax is resistant to all of them; hence rust must be controlled by growing varieties that are resistant to the races that are prevalent in a given region at a given time.

Pasmo was introduced into the United States from Argentina about 1915, but it did not become destructive in this country until about 30 years later. At present no varieties of flax are resistant to this disease, but varieties differ considerably in their susceptibility to it.

Damage of various kinds to seed bolls is common and is induced by many agents, such as fungi, viruses, insects, and hail (see INSECTA). This damage is often associated with infected or shriveled seed (Fig. 4). Damage to the seed coat during harvest is common, especially when the seed is dry; sometimes more than 75% of the seed may be cracked. These cracks, although frequently microscopic in size, serve as avenues of entrance for many species of fungi. Seed rot and seedling blight are usually associated with damaged seed (Fig. 5). Stands usually are greatly improved if damaged seed are treated with an appropriate fungicidal chemical (see FUNGISTAT AND FUNGICIDE).

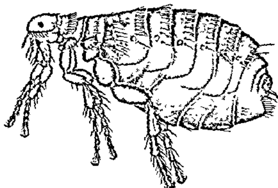
Most of the flax diseases can be controlled, or materially reduced, by planting sound and disease-free seed, treating seed with a suitable fungicide, using good cultural practices including early planting, practicing sanitation and a good cropping sequence, and growing disease-resistant varieties. See PLANT DISEASE CONTROL. [J.J.C.]

Bibliography: See AGRICULTURAL SCIENCE (PLANT); PLANT DISEASE.

Flea

Any member of the insect order Siphonaptera. Fleas are wingless, laterally compressed, small, ectoparasitic insects. They have piercing-sucking mouthparts, complete metamorphosis, and their legs are modified for jumping.

There are about 1100 species of fleas, all of which are parasitic on warm-blooded vertebrates. Fleas are important in spite of their small size because they are vectors for organisms that cause disease, some are intermediate hosts for tapeworms, at least two of which infest man; another burrows under the skin of man. Their bites can be very irritating and sometimes fatal to domestic and wild animals.



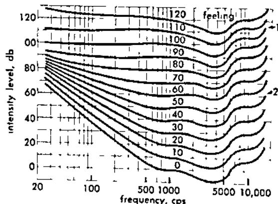
The dog flea, *Ctenocephalus canis*; length about 1/10 in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

The most important pathogen carried by fleas causes bubonic plague (see PLAGUE). This bacterial disease is primarily a disease of rodents but can be transmitted to man through the bite of a flea after the latter has fed on an infected rodent. A mild form of typhus, not to be confused with the severe type transmitted by lice, is also carried by fleas. The tropical chigger, or chigoe, is a flea which burrows into the skin and under the nails and claws of men and other mammals (see CHIGGER).

All fleas are small, the largest, about 5 mm long, being a California species, *Hystricopsylla gigas*, which lives on various native rodents. *Pulex irritans*, the human flea, which also lives on rats, appears to be a native of Europe, but it is now widely distributed. It is an effective plague carrier and a common human parasite, especially in Europe, but it is also found in America. Like most other fleas it has a highly variable life history, requiring as little as 2 weeks for a complete generation in some areas but up to 11 weeks in others. This species is said to live 4 months without feeding, and 24 months if fed occasionally. See ENTOMOLOGY; SIPHONAPTERA. [J.D.B.]

Fletcher-Munson contours

Curves showing equal loudness levels for pure tones, plotted on a graph of frequency versus sound intensity. The Fletcher-Munson contours, also called equal-loudness contours, are shown in the figure. They were first published in a paper by H. Fletcher and W. A. Munson, and represent rather extensive experimental data taken at the Bell Telephone laboratories. They were obtained by judgment tests taken by a number of individuals. The observer determined when two tones of different frequencies sounded equally loud to him. The averages of a large number of observations are shown by these curves. The ordinate gives intensity level, which is the number of decibels (db) th



Fletcher-Munson contours.

sound is above zero intensity level, the latter corresponding to a sound wave which carries a power of 10^{-16} watts/cm^2 in its wavefront The abscissa gives the frequency of the tone.

le
lo

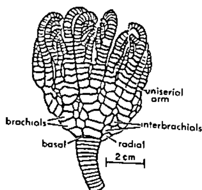
quencies below 1000 cycles per second (cps) have approximately the same loudness when they have an intensity level of 100 db, as shown by Curve 1. For lower levels, however, this is not true. A tone of 1000 cycles and intensity level of 40 db will sound as loud as a 100-cycle tone of 62-db intensity level (Curve 2).

These curves show why a listener who is near a band hears the bass much more prominently than when he is some distance away from it. Modern hi-fi experts make use of these contours in the design of amplifiers for reproducing systems. At low levels of reproduction, the low-frequency components of the sound may be enhanced in accordance with these curves to preserve the effect of good bass. *See* HEARING; LOUDNESS. [H.FL.]

Bibliography: H. Fletcher, *Speech and Hearing in Communication*, 2d ed., 1953.

Flexibilia

An extinct subclass of stalked or creeping Crinoidea which includes some 50 Paleozoic genera, ranging from Ordovician to Permian times. A flexible tegmen was present with open ambulacral grooves. One of the 5 oral plates served as a madreporite. The anus was at the tip of a short siphon About 3 of the lower brachial ossicles of each arm were loosely incorporated into the dicyclic calyx. The 3 infrabasals were sometimes obscured by the stem, but the 5 basals and 5 radials remained conspicuous and movably articulated. The uniserial arms lacked pinnules but branched freely, arching inward to form a globular crown. The cylindrical stem lacked cirri, and was sometimes discarded, as in *Edriocrinus*, to produce a creeping adult. *See* CRINOIDEA. [H B F.]



Synerocrinus, Flexibilia. (*After F. Bather, 1900*)

Flicker

Any of five different woodpeckers of the genus *Colaptes*, two species of which occur in the United States. *C. auratus*, represented by several subspecies, is known in the East as the yellow-shafted flicker and in the West as the red-shafted flicker.



The flicker, *Colaptes auratus*; length to 13 in. (*From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949*)

The principal difference between the two is the color of the undersides of the wings and tail. The flickers are ant-eating birds and spend much time in the summer hunting ants from their nests and runways. Flickers are the only brown-backed woodpeckers and are further characterized by a prominent white patch on the rump. *See* PICIFORMES; WOODPECKER. [J D B]

Flight

Traditionally, flight has dealt with motion in or through the air, but flight now includes motion at distances so far from the earth's surface that the air has negligible effect. Natural animal flight is necessarily restricted to the denser portion of the earth's atmosphere. Artificial flight can either use

the air (*see* AERONAUTICS) or be independent of it (*see* ASTRONAUTICS).

The ballistic flight of a bullet or missile is a consequence of energy imparted to it during its ejection from a gun barrel or during the burning period of its engine. At the opposite extreme is the soaring flight of a glider or a hawk that depends on the energy from local rising air currents. True natural flight, as for birds, bats, and insects, and true artificial flight, as for men in aircraft, requires the continuous expenditure of energy to achieve fully controlled progress through the air. Historically, insects flew first, followed by reptile groups that are now extinct, then birds, bats, and finally man.

The greater the size of an animal, the greater is the problem of remaining airborne, because, although the weight increases as the cube of the linear dimensions, the surface area increases only as the square. The practical limit seems to have been reached in the largest birds. Probably this is also why the insects were the first and biologically speaking the most successful fliers, because they are the smallest. Gossamer spiders, some mites, and minute insects are able to remain airborne although they have no wings.

Wing motion. In all flying animals, both lift and thrust are provided by the predominantly up and down movement of a pair of bilaterally extended planes. In the more primitive insects, there are two pairs of such planes which may be partially independent.

The wings of insects are merely double extensions of the body wall supported by characteristic thickenings known as veins, especially near the leading edges (*see* INSECT PHYSIOLOGY). Those of birds are built of feathers which are tattered exaggerated reptilian scales, outgrowths of the skin, supported by the whole skeletal structure of the front leg or arm. Those of the bats are built of the skin itself, supported principally by the fingers of the front limb; a sort of ultimate glorification of the webbed foot of a duck or frog.

The movements of animal wings in level flight usually show a forward component on the downstroke and a backward component on the upstroke, either relative to the body axis or to the vertical or both (Fig. 1). This has the effect of giving a downward component to the main direction of the slipstream and hence contributes lift. In hovering flight, this is usually tremendously exaggerated. On the downward and forward stroke, the wing is twisted so that the leading edge is below the trailing edge, while on the upward and backward stroke the reverse is true. This twisting is accomplished principally by two interactions: that between the complex articulations at the base of the wing and the muscular forces acting on it, and that between the aerodynamic forces and the strength distribution of the wing structure itself. Special muscles other than those which power the wing stroke may control the twisting. In aeronautical terms, this twisting may be regarded as pitch reversal in an air-

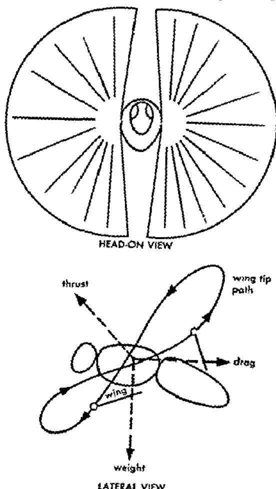


Fig. 1. Typical movement of animal wing in flight.

screw, accompanying a reversal in direction of rotation (*see* HELICOPTER). To take this analogy further, most flying animals can be regarded as aircraft in which the functions of wings and airscrews are combined in a pair of single-bladed, semi-rotary, reversible pitch rotors, mounted near the center of gravity, vibrating at close to their natural mechanical frequency (Fig. 2). *See* ORNITHOPTER.

In some of the smaller insect species, wing beat frequencies up to 1000 cps have been recorded. The amplitude of wing movements is greater than commonly supposed, and in many insects may exceed 180°.

In detail, wing movement is more complex; much remains to be learned about the significance of all the details of wing structure and above all of the subtleties of control required in the execution of complex aerial maneuvers.

Power is applied to the wings by large, highly specialized muscles: in insects, the longitudinal median dorsal and the tergosternal muscles of the thorax, acting indirectly through an elaborate system of elastic plates and levers; in birds and bats, the pectoral muscles with direct pulls from the

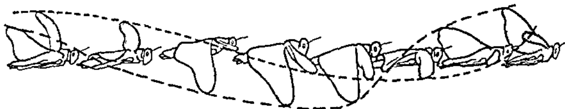


Fig. 2 Example of combined twisting and flapping of insect wings in flight.

limb bones to a tremendously enlarged keel on the breast bone. The primary fuels providing energy to the muscles are probably always either carbohydrates or fats.

Flight capabilities. Many wild guesses at performance figures have been published, especially for insects. Probably the maximum air speed of which any living insect is capable in level flight is about 36 mph; this would be possible only in a short burst. Continuous flight is limited probably to 24 mph. Both of these speeds are approached only by the larger insects such as hawk moths and dragonflies. The larger an animal, the higher is its maximum speed likely to be, other things being equal. It is not surprising therefore, that birds have been recorded flying at speeds up to 60 mph for considerable distances, and approaching 100 mph for shorter distances or in a dive.

If speed has often been overestimated, range is commonly underestimated. The monarch butterfly is capable of traveling 650 air mi without feeding, and even such fragile insects as mosquitoes may travel 33 mi, also without feeding. Hummingbirds, relatively short-range species, have been recorded to cross the Gulf of Mexico. Other birds can do much better; flights of 300–400 mi in 24 hours are apparently commonplace, and flights of over 700 mi have been recorded. Relatively small species of bats have been reported to return to their home caves from distances over 200 mi; speeds up to 20 mph seem normal.

Only the birds have learned to make significant use of air currents in gliding flight; perhaps only they have needed to. [B.HO.]

Bibliography: M. Eisentraut, Die deutschen Fledermause, eine biologische Studie, *Monographie der Wildsäugetiere*, No. 2, Leipzig, 1937; J. W. S. Pringle, *Insect Flight*, 1957; J. H. Storer, *The Flight of Birds Analyzed through Slow-Motion Photography*, 1948.

Flight characteristics

The behavior of an aircraft while airborne. A vehicle on the ground is constrained in its motions by the surface of the ground, but an airborne system is free to move in any direction. For example, it may roll upside down or slip sideways, as well as move along any desired flight path.

Vehicular requirements. Any airborne system must meet at least the following requirements. It must have the capability of sustaining itself in the air under all conditions that it is expected to encounter during its missions. Its structure must withstand all reasonably probable loads without suffering permanent deformation. Its propulsive system must enable it to achieve its limits of speed and range. It must incorporate a safe, logical, and effective method for the control of its trajectory in the air. It must possess safe means for leaving and returning to the ground, water, or parent craft.

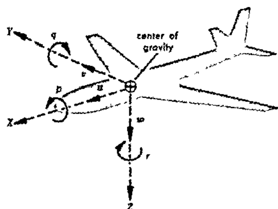
In developing an aircraft to meet these requirements, as specified for a particular mission or set of missions, its flight characteristics are described in terms of over-all performance and of handling characteristics. The separation of flight characteristics into the two broad categories of performance and handling is one of convenience rather than of fact. An airplane in flight is a dynamic system whose behavior depends on all its characteristics interacting with each other, rather than on an isolated subset of these characteristics.

Again, as a matter of convenience, performance and handling qualities may both be broken into static and dynamic groupings, resulting in a set of four related headings that categorize aircraft flight characteristics. These headings are (1) static performance, (2) dynamic performance, (3) static handling qualities, and (4) dynamic handling qualities. Static as used here has the Newtonian connotation that statics and dynamics differ only in that dynamic motion involves accelerations of various types whereas static motion involves only constant velocities.

Low- and high-speed aircraft. When dealing with low-speed aircraft (airplanes which are totally subsonic), aircraft performance may be accurately treated as a static or quasistatic phenomenon; however, the approximations incurred in using this approach become far too inappropriate when dealing with supersonic vehicles. The performance of these vehicles must be analyzed by nonconservative particle dynamics. To evaluate maximum performance capabilities, the procedures of the calculus of variations must be used.

In an analogous way, the handling qualities of low-speed aircraft may be approximately treated by static techniques, and proper control response may be provided altogether by direct aerodynamic means. On the other hand, high-speed aircraft generally require the synthesis of handling characteristics and the creation of apparent fundamental stability by means of optimum adaptive feedback servo loops. Actual control moments on the aircraft at high speeds may be produced by aerodynamic means alone or in combination with reaction units which directly produce a moment about the

airplane's center of gravity. At high speeds the ability of an airplane to turn rapidly or to zoom to its maximum dynamic ceiling depends strongly on the available control-system authority and upon the ability of the airframe to develop substantial normal load factors when these are required to fulfill trajectory demands.



Axis	Force along	Moment about	Linear Velocity	Angular Displacement	Angular Velocity	Inertia
X	F_x	L	u	ϕ	p	I_x
Y	F_y	M	v	θ	q	I_y
Z	F_z	N	w	ψ	r	I_z

Fig. 1. Airplane can move along or rotate about each of three axes.

Degrees of freedom. An airplane treated as a rigid system requires a minimum of six equations of motion to define its state of being since it has three translational and three rotational degrees of freedom (Fig. 1) Because aerodynamic forces are conventionally measured with respect to an axis system attached to the hypothetical relative wind vector, the airplane's equations of motion are frequently written with respect to this wind-axis system, however, rotational inertia effects depend on the change in inclination of an axis system attached to the airplane, and this system, known as the body-axis system, must be used in conjunction with the wind-axis system to define completely the airplane's motion as a rigid body by reference to a fixed- or moving-gravity axis system (Fig. 2).

Equations of motion. To write the equations of motion in precise accord with the Newtonian concept of space, one would first have to establish a so-called fixed point in space and consider rectilinear and rotary motion with respect to that point as origin; however, for speeds up to about 20,000 mph reasonable accuracy is achieved by considering the center of the earth to be a fixed point. At speeds of 3000 mph or less the curvature of the earth can be ignored and the surface of the earth treated as a plane containing the fixed point with respect to which motion takes place.

The general equations for the motion of a particle moving entirely within a navigational or terrestrial great circle relative to the rotating earth

wind Z or lift axis
wind X or drag axis
wind Y or side force axis
center of gravity
body longitudinal or rolling axis
L', p, ϕ
X, x
N, r, ψ
Y, y M, q', Θ
body lateral or pitching axis
body vertical or yawning axis
Z,z
relative wind vector
angle of attack
side slip angle
local horizontal or gravity axis X

Airplane Axis Conventions

L', M, N are moments about the body axes
p, q', r are angular rates about the body axes
ϕ, Θ, ψ are angular displacements of the body axes

Fig 2. Conventional wind and body axes systems together with the conventions of angular measurement.

with fixed point of origin at the earth's center are

$$F_{x_i} = -m\dot{V}_r + m\Omega^2 h_a\{\sin^2\phi_1 \sin\gamma - \cos^2\phi_1 \cos\Theta_1 \sin(\Theta_1 - \gamma)\}$$

$$F_{y_i} = -2\Omega m V_r \cos\phi_1 \sin(\Theta_1 - \gamma) - m\Omega^2 h_a \sin\phi_1 \cos\phi_1 \sin\Theta_1 + mV_r\dot{\phi}_1\{\cos(\Theta_1 - \gamma) + \tan\Theta_1 \sin(\Theta_1 - \gamma)\} + m\Omega h_a\dot{\phi}_1 \sin\phi_1 \tan\Theta_1 \sin\Theta_1$$

$$F_{z_i} = mV_r(\dot{\Theta}_1 - \dot{\gamma}) + 2mV_r\Omega \sin\phi_1 + m\Omega^2 h_a\{\sin^2\phi_1 \cos\gamma + \cos^2\phi_1 \cos\Theta_1 \cos(\Theta_1 - \gamma)\}$$

where F_{x_i}, F_{y_i}, and F_{z_i} are the inertial forces along the x, y, and z axes, V_r is the relative wind velocity, ϕ_1 is the angle between the instantaneous plane of motion of V_r and the polar axis, Θ_1 is the angular location of the particle in the plane of motion measured from the Equator, m is the vehicle mass, γ is the inclination of V_r to the local horizontal, Ω is the rotative speed of the earth, and $(\cdot)$ is total time derivative.

For speeds of less than 3000 mph, the quantity $\dot{\Theta}_1 = V_r \cos\gamma/h_a$ is small and may be set to zero; moreover Ω produces no sizeable effects, and terms in Ω may be dropped. Under these circumstances the set of equations reduces to

$$F_{x_i} = -m\dot{V}_r \qquad F_{y_i} = 0 \qquad F_{z_i} = -mV_r\dot{\gamma} \quad (1)$$

where F_{y_i} is zero due to the use of the wind-axis reference system.

The complete planar particle equations of motion are obtained by balancing *inertial* forces with direct aerodynamic and thrust *forces* produced by the airframe. Thus, if F_e is thrust, L and D are the trimmed lift and drag forces acting on the airplane, and α_T is thrust line inclination, equilibrium gives

$$F_e \cos\alpha_T - m\dot{V}_r - D - mg\sin\gamma = 0$$
$$F_e \sin\alpha_T - mV_r\dot{\gamma} - mg\cos\gamma + L = 0 \quad (2)$$

Trimmed, as used here to modify *lift* and drag, means that these are the forces which exist when the control surfaces have been deflected to achieve the required lift and moment balance at each point along the flight path. The *trimmed* forces are closely approximated by those *obtained* by simply balancing out the static stability moment of the aircraft.

For low-speed airplanes, a static performance analysis can be made which ignores accelerations and thrust-line angle of attack. In such a case the governing equations of motion are assumed to be

$$F_e - D - mg\sin\gamma = 0 \qquad L - mg\cos\gamma = 0 \quad (3)$$

Static characteristics. If γ is small, lift and weight are approximately equal so that $L \approx mg \approx W$.

Rate of climb. When approximations such as these are permissible, the static rate of climb dh/dt is found directly from Eq. (3) to be

$$dh/dt = V_r \sin\gamma = V_r(F_e - D)/W \quad (4)$$

Because V_rF_e is the power *available* and V_rD is the power required *to fly* level, the quantity $V_r(F_e - D)$ is designated as *excess* power, and Eq. (4) stipulates that maximum static rate of climb is obtained at the speed where the excess power is greatest. Similarly from Eq. (3)

$$\sin\gamma = \frac{F_e - D}{W} \quad (5)$$

which states that the maximum static climb angle is obtained at the speed producing the greatest excess thrust.

The static time to climb is obtained from the integral

$$t = \int_0^h \frac{dh}{V_r \sin\gamma} = \int_0^h \frac{W\,dh}{V_r(F_e - D)} \quad (6)$$

where h is geometric height.

Ceiling. The maximum static altitude or *static ceiling* of an airplane is that altitude at which the rate of climb defined by Eq. (4) falls to zero, and the maximum static speed of an airplane at any altitude is the speed where $F_e - D$ becomes zero under increasing speed conditions. The minimum level-flight speed is limited either by aerodynamic stall or by the condition of $F_e - D \to 0$, as speed is decreased. At the absolute ceiling the minimum and maximum speed are the same. Flight under [illegible] b [illegible] o achie[illegible] the maxi-mum [illegible] tude [illegible] possible. This ceiling is called the service ceiling of an airplane.

Range. To complete the description of static performance, range R and endurance E under quasi-static conditions in level flights are obtained by maximizing the integrals

$$R = \int dR \qquad E = \int dE$$

For a propeller-driven airplane

$$dR = -V_r\,dW/cP$$

where c is brake *specific fuel* consumption, P is power, and W is weight.

The power setting is determined as the power required for steady level flight with a propeller efficiency η, or $P = DV_r/\eta$; thus $dR = \eta\,dW/cD$. If L equals W, then multiplication of dR by L/W does not change dR and

$$dR = (\eta/c)(L/D)(dW/W)$$

If η and c are constant, then maximum range is obtained by flying in such a way that L/D is maximum.

In coefficient form $L = qC_LS$, $D = qC_DS$, where $q = \rho V_r^2/2$ is the dynamic pressure, ρ being the ambient density, S is wing area, C_L is lift coefficient, and C_D is drag coefficient. Thus

$$dR = \frac{\eta}{c}\frac{C_L}{C_D}\frac{dW}{W} \quad (7)$$

If conversion factors are applied so that R is ex-

pressed in statute miles and c in lbs/bhp-hr, integration of Eq. (7) gives

$$R = 375 \frac{\eta}{c} \frac{C_L}{C_D} \ln \left[\frac{\text{initial weight}}{\text{final weight}}\right] \quad (8)$$

Equation (8) is the classic Brequet equation for range of propeller-driven aircraft and gives rise to the statement that maximum range for such aircraft is approximately obtained by flying at the best lift-drag ratio. Because η and c are not really constants, Eq. (8) can, in some instances, be misleading, and for precise analysis stepwise range integration is required.

Endurance. The time of flight is given as $dt = dW/cP$ where t is time in seconds, and the endurance equation for piston-engined aircraft with constant η and c is

$$E = \int dt = \frac{\eta}{c} \frac{C_L^{3/2}}{C_D} \sqrt{2\rho s} \times \left(\frac{1}{\sqrt{\text{final weight}}} - \frac{1}{\sqrt{\text{initial weight}}}\right) \quad (9)$$

This equation gives rise to the approximate criterion that maximum endurance flight occurs when operation is at the maximum value of the ratio $C_L^{3/2}/C_D$

Jet range and endurance. These relations do not hold for jet aircraft because for these vehicles the rate of fuel consumption is approximately proportional to the thrust developed rather than to the power. Thus, for a pure jet airplane

$$R = \int \frac{V_r \, dW}{c_t T} \quad (10)$$

where c_t is thrust specific fuel consumption and T is thrust. Similarly

$$E = \int \frac{dW}{c_t T} \quad (11)$$

For R in statute miles, time in seconds, c_t in lb/lb of thrust per hr, and for the case of constant altitude and constant c_t, Eq. (10) and (11) integrate to

$$R = \frac{1\,929}{c_t \sqrt{\rho s}} \frac{C_L^{1/2}}{C_D} (\sqrt{\text{initial weight}} - \sqrt{\text{final weight}}) \quad (12)$$

$$E = \frac{1}{c_t} \frac{C_L}{C_D} \ln \left(\frac{\text{initial weight}}{\text{final weight}}\right) \quad (13)$$

Equations (12) and (13) give rise to the approximate criteria that best jet range occurs when flight is conducted at $C_L^{1/2}/C_D$ maximum, and best endurance results from flying at C_L/C_D maximum. As in the case of propeller-driven aircraft, these results are quite approximate, and c_t for instance varies with both speed and altitude. Operationally a jet airplane obtains better range during a cruise climb than it does at constant altitude. In this case, a fixed throttle setting is used which at any given initial altitude maximizes the quantity

$$\frac{C_L^{1/2}/C_D}{c_t \sqrt{\rho S}}$$

and Mach number and angle of attack are maintained constant thereafter, which leads to a gradual climb as fuel is consumed and the weight decreases. In any event, a jet will not have its most efficient cruise at a Mach number greater than drag divergence, so that the limiting cruise Mach number cannot exceed the Mach number at which ratio $C_L^{1/2}/C_D$ falls off because of too rapid an increase of drag coefficient with speed.

Glide and dive. Static analysis may also be used to define gliding and diving performance wherein equilibrium is presumed to exist. If equilibrium can be established

$$\tan \gamma = \text{glide angle} = -D/L = -\frac{1}{C_L/C_D} \quad (14)$$

Similarly, terminal static Mach number M_T in a dive directly toward the earth is given by

$$M_T = \sqrt{\frac{2w}{\gamma p S C_{D\,\min}} \cdot \frac{C_{D\,\min}}{C_D}} \quad (15)$$

where p is ambient pressure in lb/ft^2, γ is specific heat ratio, which for air is 1.4, $C_{D\,\min}$ is minimum drag coefficient of airplane at the most favorable Mach number, and $C_{D\,\min}/C_D$ is the ratio of minimum drag coefficient to actual drag coefficient at M_T. Because $C_D = f(M_T)$, Eq. (15) is normally solved iteratively.

Dynamic characteristics. Whereas the static presentation serves to introduce fundamental concepts as well as terminology, the precision of static analyses is such that if used to analyze high-performance aircraft they will produce extremely large errors. For such vehicles, inertial forces are no longer small in comparison to lift, drag, and thrust, so that the assumptions of the static analyses are invalid.

Performance. Normally, determination of dynamic performance and stability for high-performance vehicles involves the integration of a set of nonlinear differential equations, and this process is practical, in general, only when using a large-scale computer. Also, because of the pseudoholonomic nature of the coupled equations of motion given as Eq. (2), closed-form solutions are not possible, because of the need for application of the least action uniqueness condition at all points along the flight path

Considering the requirement for minimum time to climb as typical, dynamic performance techniques call for maximizing the integral

$$h = \int_0^t \frac{dh}{dt} \, dt \quad (16)$$

between selected end conditions of V_T and γ at $t = 0$, and $t = T$. Unlike a static analysis, there exists a different best path for each set of end conditions, and the concepts of static-absolute and

static-service ceilings no longer apply. Indeed, using dynamic techniques, a Mach 3 fighter can zoom to an altitude as much as 50,000 ft above its static ceiling. The general procedure used to maximize an integral is to employ the Euler equations of variational calculus, after first restraining the variational integrand in accord with the requirement that the equations of motion shall always be satisfied.

In particular, for the climb problem $dh/dt = V_r \sin\gamma$ but explicity from the first of Eq (2)

$$\frac{mg\sin\gamma}{F_e\cos\alpha_T - m\dot{V}_r - D} = 1 \tag{17}$$

at all points along the flight path, whereas implicity because D is composed of parasite and induced drag, and induced drag depends on C_L to an exponent greater than 2, the second of Eq. (2) provides a pseudoholonomic relation between $\dot{V}_r$ and $\dot{\gamma}$; that is, there exist two or more real values of $\dot{\gamma}$ at each point which satisfy the equations of motion when V_r, γ and $\dot{V}_r$ are specified

One form of restrained integrand, therefore, becomes

$$G = V_r\sin\gamma + \frac{\lambda mg\sin\gamma}{F_e\cos\alpha_T - m\dot{V}_r - D} \tag{18}$$

where λ is a constant Lagrange multiplier for the case shown. On the basis of this definition of G, the Euler relations are

$$\frac{d}{dt}\left(\frac{\partial G}{\partial\dot{\gamma}}\right) - \frac{\partial G}{\partial\gamma} = 0$$
$$\frac{d}{dt}\left(\frac{\partial G}{\partial\dot{V}_r}\right) - \frac{\partial G}{\partial V_r} = 0 \tag{19}$$

These equations in turn produce a nonlinear differential equation in $\ddot{V}_r$, which, together with the equations of motion and the least action uniqueness condition, serves to define the optimum climb flight path completely. The equation set is readily solved on a digital computer when it is provided with a mathematical description of atmosphere, engine, and airframe characteristics.

The nature of minimum time turns, fixed- or programmed-throttle maximum range paths, and like maneuvers may all be determined using the indicated variational techniques.

by feedback-servo stability augmentation including such devices as command filters and self-adaptive optimum controllers. The analysis of dynamic response thus involves the study of systems which are generally nonlinear and whose over-all characteristics are the result of the cascading of several servo elements with the actual airframe dynamics. As in the study of dynamic performance, the nonlinearity associated with the dynamic response generally requires the use of digital computing equipment to achieve even reasonably accurate answers to real problems.

To complete the picture of dynamic characteristics, the problems of gaining and losing speed during take-off and landing respectively, must be mentioned. Because either of these problems involves a single degree of freedom, analysis is readily handled starting with the generalized distance definition obtained from Newton's second law. Thus, for ds the increment of take-off or landing distance

$$ds = \frac{dV_{as}^2}{2a} - V_w\,dt \tag{20}$$

where V_{as} is air speed, V_w is constant wind speed, and a is net acceleration acting on system. If F_m is the mean accelerating force acting during take-off (the average value of thrust-drag-rolling resistance), then

$$S = \frac{W}{2gF_m}(V_e - V_w)^2 \tag{21}$$

where W is weight of aircraft in lb, and V_e is air speed at take-off point in ft/sec. Normally V_e = 1.2, which is the stalling speed of take-off configuration.

The landing roll distance is given by Eq. (21) with V_g being the ground-contact speed and F_m the mean applied decelerating force.

The transition between take-off and climb-out is normally accomplished at constant or increasing air speed, and should never be attempted at load factors great enough to produce a decrease in speed, which would tend to produce aerodynamic stall. On the other hand, the transition between approach and landing should be accomplished so as to provide as low a speed as possible on ground contact with the restriction that stall must not be approached so closely that loss of control is encountered. *See* AIRCRAFT TESTING. [D.O.D.]

Bibliography: D. O. Dommasch, S. S. Sherby and T. F. Connolly, *Airplane Aerodynamics*, 2d ed., 1957; C. D. Perkins and R. E. Hage, *Airplane Performance, Stability and Control*, 1949.

Flight controls

The control systems that govern aircraft attitude and direction of motion. Three forms of flight controls are (1) primary flight controls, (2) auxiliary flight controls, and (3) automatic controls. Primary controls consist of the longitudinal or pitch attitude controls (elevators), the lateral or roll attitude controls (ailerons), and the directional or yaw attitude controls (rudder). A primary flight control found on helicopters and other vertical-rising machines is the lift control. On helicopters this is the collective pitch control. Auxiliary flight controls include trim for the primary flight controls, landing flap, leading-edge flap or slat, adjustable wing- or tail-surface incidence, adjustable wing sweepback, and speed- or dive-brake controls. Automatic flight-control systems (AFCS) supple-

ment or replace the human pilot as a source of guidance, stabilization, and actuating power. In a general sense, flight controls are a class of servomechanisms or automatic feedback systems in which aircraft control is achieved by computation and amplification performed on low-power level signals generated by a human pilot or automatic guidance system.

Primary flight controls. Primary flight controls have a set of manually operated levers, wheels, and pedals, called the cockpit controls, located in front of the pilot. The motions or forces applied to the cockpit controls are transmitted by cables, push-pull rods, and in some cases by electrical signals to the control surfaces or devices that apply the forces governing aircraft attitude and direction of motion. A simplified schematic diagram of a primary flight control system is presented as Fig. 1.

Cockpit controls. Cockpit controls are similar in all classes and sizes of aircraft because their dimensions, travels, and actuating-force levels are determined by human limitations. Longitudinal motion or force of the cockpit controller governs pitch attitude in a natural sense; that is, forward motion or force corresponds to nose down and rearward motion or force corresponds to nose up. This controller may be a lever (control column), pivoted or sliding in front of the pilot, or a smaller lever located to his right (side stick). Lateral motion of the cockpit controller governs roll attitude in a natural sense; that is, controller motion to the right, or clockwise, generates right wing down or right turn, and motion to the left generates left wing down or left turn. The lateral controller, generally a part of the longitudinal controller, often resembles an automobile steering wheel for centerline installations (control wheel). The directional cockpit controller, when provided, is traditionally a foot control (rudder pedals). Depressing the right pedal causes nose-right aircraft motion; depressing the left pedal causes nose-left motion. This is not an instinctive controller, because it operates in a sense opposite to that of bicycle handlebars. The lift or collective-pitch cockpit flight controller is a separate lever, operating in an instinctive manner. Motion or force upward causes increased lift and an upward path displacement.

Cockpit controls for auxiliary flight-control systems and automatic flight-control systems are not standardized with respect to size or location. Directions of motion or applied forces correspond to the primary flight controls, or are instinctive.

Transmission and amplification. The low-power-level signals applied to cockpit controls by the human pilot are transmitted and on large or fast aircraft are amplified by a series of mechanical, hydraulic, and electrical components. The relation-

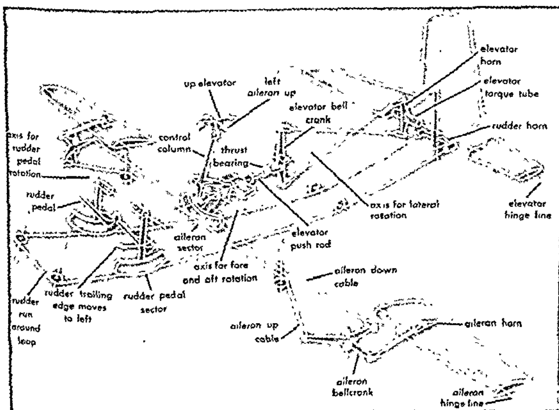


Fig. 1. Simplified diagram of typical primary flight controls. (C. H. Chatfield, C. F. Taylor, and S. Ober, *The Airplane and Its Engine*, 5th ed., McGraw-Hill, 1949)

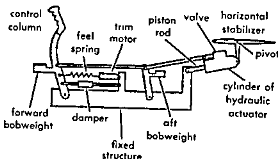


Fig. 2. Schematic diagram of longitudinal control system employing irreversible hydraulic actuator.

ship governing transmission and amplification for simple flight-control systems is an application of the virtual work principle, written as

$$F = (\partial\delta/\partial x)H/B$$

where F is the force applied to the cockpit controller, $\partial\delta/\partial x$ is the gearing ratio, or rate of change of control surface angular travel with displacement of the point of application of control force, H is the torque or hinge-moment reaction about the control-surface hinge line, and B is the boost or amplification ratio. Using consistent units of pounds, feet, and radians, the gearing ratio is of the order of 1 0 radian/ft The boost ratio may vary from about 2 to infinity (corresponding to irreversibility). The control hinge moment H may be several thousand pound-feet, in order to control effectively a large or fast aircraft. Impossibly large values of cockpit controller force F are avoided by aerodynamic balancing, with which the hinge moment is reduced to controllable levels by means of aerodynamic forces on leading-edge balances or tabs (*see* ELEVATOR, AIRCRAFT). Large hinge moments are also developed by internally powered electric or hydraulic actuators. Until the advent of supersonic flight in 1947, aerodynamic balancing schemes were adequate for all classes and sizes of aircraft. However, supersonic airplanes and guided missiles must employ powerful actuators to move their control surfaces. A schematic diagram of a longitudinal control system employing a hydraulic actuator is shown in Fig. 2. The actuator in Fig. 2 is an irreversible closed-loop hydraulic servomechanism. The boost ratio B is infinite, and pilot control forces are generated by the spring, damper, and bobweight, constituting the artificial feel system.

Automatic flight controls. All functions performed by human pilots in the control of aircraft can be performed automatically, using as basic inputs either an arbitrary preset program or the same sensory inputs as those available to human beings. Automatic flight-control systems are used to increase precision of control, and to relieve the human pilot of routine tasks. A block diagram of a typical automatic flight-control system, for performing automatic approaches to a landing field, is presented in Fig. 3. The distance of the aircraft from the desired landing approach path is detected by radio receivers and resolved into lateral and vertical errors. A computation is made in the approach coupler for the optimum aircraft bank angle, and hence turning rate, as a function of lateral error and its derivative. The output of the approach coupler, the desired bank angle, is compared with the actual bank angle as measured by the vertical gyroscope, and the difference, or bank error, is a driving signal to the lateral control or aileron servo channel. Synthesis and analysis of such a system for rapid nonoscillatory response are made by the techniques of servomechanism analysis (*see* SERVOMECHANISM). The components of automatic flight control systems generally include the devices described below.

Sensing devices. Precision instruments sense the aircraft attitude and direction of motion and generate corresponding signals. These sensing devices may be gyroscopes, accelerometers, wind vanes, barometric devices, and radio, radar, infrared, or photoelectric receivers.

Computers. High-speed special-purpose computers develop steering signals at low power levels, based on sensing device outputs, guidance signals from the human pilot, and predetermined programs.

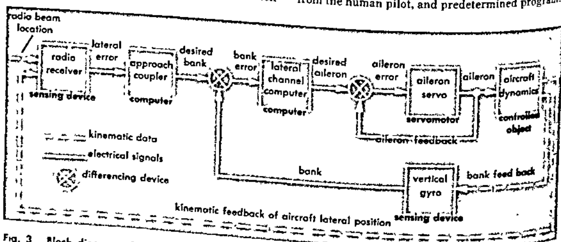


Fig. 3 Block diagram of automatic flight-control system, for landing approaches (lateral channel only).

Both analog and digital computers are used in automatic flight-control systems. Frequency and time-sharing techniques may be used to make these computers available for tasks other than flight control.

Servomotors. The large power levels required to move aerodynamic, reaction, or inertial control devices rapidly and precisely are provided by servomotors. Hydraulic, electric, pneumatic, or hot-gas servos are used. The development of hydraulic valves that can be actuated by either low-energy electrical signals or mechanical inputs from the cockpit controllers has led to the integrated flight-control system. A single servomotor actuates the control surfaces or devices when the aircraft is controlled by either the human pilot or the automatic system, with reduced complexity.

Adaptive, redundant, and self-monitoring systems Adaptive automatic flight-control systems vary their parameters to maintain optimum performance under all flight conditions. The five main types are (1) passive adaptation, (2) input-signal adaptation, (3) extremum adaptation, (4) system-variable adaptation, and (5) system-characteristic adaptation.

Many automatic flight-control systems are capable of destroying the aircraft as a result of failures in their components. High reliability is sought by redundancy and rugged design. Self-monitoring automatic flight-control systems check system performance against indications from independent references. See AIRPLANE; RUDDER, AIRCRAFT.

[M.J.AB.]

Bibliography: H. Templeton, *Massbalancing of Aircraft Control Surfaces*, 1954; J. G. Truxal, *Automatic Feedback Control System Synthesis*, 1955; U.S. Air Force Air Research and Development Command, *Handbook of Instructions for Aircraft Designers*, vol. 1, 1955; U.S. Bureau of Aeronautics, *Fundamentals of Design of Piloted Aircraft Flight Control Systems*, 6 vols., 1952

Flight dynamics

The study of the motion of an aircraft or missile. Flight dynamics is generally concerned with the transient or short-term effects, which have to do with the stability and control of the vehicle, rather than with calculating such performance as ultimate range, altitude, or velocity of the vehicle (*see* FLIGHT CHARACTERISTICS). Sometimes, however, the two are treated together, particularly in the case of aircraft.

A typical problem in flight dynamics considers aerodynamic forces and the thrust of an engine (rocket or other) on a flexible airframe with a control system of specified characteristics interposed (*see* AEROELASTICITY; WING STRUCTURE). It may then be desired to find the effect on the whole vehicle of a sudden change in attitude caused by a gust, or to find the effect of changing the flight speed on the controllability of the vehicle. Vehicles which often have satisfactory control characteristics at one speed may become uncontrollable at another because of a shift in the aerodynamic loads. A problem in flight dynamics then typically involves aerodynamics intermingled with the characteristics of the vehicle structure and its propulsion system. To this is added the control system with its own peculiarities and responses which must be tailored to all the rest. Many problems of this sort are now solved with the use of large-scale analog or digital computers. [J.R.SE.]

Flight science

The sum total of all knowledge that enables man to accomplish flight. Flight science is compounded of both science and engineering. It is concerned with airplanes, missiles, and manned and unmanned space vehicles.

The scope of flight science is illustrated by some of the diverse fields which are included within it, such as electronics, aerodynamics, propulsion engineering, structural engineering, nuclear engineering, metallurgy, chemistry, space medicine, certain parts of astronomy, mathematics, classical and modern physics, and other branches of engineering such as civil engineering for the planning and construction of airports. These constitute the major branches of science and technology which are necessary for the solution of the problems of modern flight.

While the above are all-encompassing categories, it is useful to consider in more detail a few specific categories to see their particular contribution. Electronics, for instance, contributes in so many ways as to be almost indispensable for modern flight. Guidance, control, communication, and instrumentation are without doubt the most important elements of many modern airborne systems. *See* AIRCRAFT INSTRUMENT PANEL; COMMUNICATIONS SYSTEMS (TRAFFIC) DESIGN; COMPUTER; CONTROL SYSTEMS; FLIGHT CONTROLS; GUIDANCE SYSTEMS.

Aerodynamics is traditionally the core of flight science; it is concerned with the flow of air at flight speeds of a few feet per second to many times the speed of sound. The density of the flow may vary from that at sea level to that at hundreds of miles in altitude. *See* AEROSTATICS; HYPERSONIC FLIGHT; REENTRY, SUBSONIC FLIGHT; SUPERAERODYNAMICS; SUPERSONIC FLIGHT; TRANSONIC FLIGHT.

Chemistry contributes by making available new materials for special applications such as special plastics for heat protection, but most importantly in the synthesis of new fuels. These fuels constitute the most significant development of modern flight science. Without these advances, weapons such as a solid-propellant ICBM would be hopelessly impractical (*see* JET VELOCITY; PROPELLANT). In space flight and in the design of space cabins, biochemistry plays a vital role; it supplies the air conditioning and will ultimately provide for the reprocessing of waste material through a controlled ecology. *See* SPACE FLIGHT.

Nuclear engineering provides the greatest pote tial for high performance or long duration fl

While the practical schemes for the utilization of nuclear power are still in the making, it requires no great foresight to see that this source of power will ultimately be used to propel flight vehicles (*see* NUCLEAR AIRCRAFT PROPULSION). Closely allied with nuclear power are various schemes for electric space propulsion, which hold promise of very high specific impulses, but they also require a large amount of power. *See* ELECTROMAGNETIC PROPULSION; SOLAR ENGINE.

Structures of great efficiency have been developed as the demands of space flight as well as conventional flight have increased. Metallurgy and chemistry have made new materials available which have been quickly incorporated into structures by the structural design engineer. *See* AERONAUTICAL ENGINEERING; SPACE TECHNOLOGY.

The problem of space flight has given rise to a new technology in the design of space vehicles other than ballistic missiles. Satellites, space stations, and instrumented probes of various types are examples of vehicles made for particular special purposes. Each mission has its own particular problems, and each mission requires a creative application of technology *See* SPACE VEHICLE.

Missiles of all sorts for military application rank with military aircraft as the field of most intensive technical activity. These missiles range from the subsonic guided missile to the ballistic rocket and are designed to work under many environmental conditions such as air-to-air, ground-to-air, air-to-ground, and so forth. Some are even launched from submerged submarines. The adapting of a missile to such conditions so that it performs safely, accurately, and reliably constitutes the main application effort as compared to space flight where a controlled launching condition including a long countdown is often possible. *See* MISSILE.

Another area of technology often overlooked among the more glamorous fields is that of ground support equipment. This important equipment often determines the success or failure of many flights. It consists of all the launching consoles, fueling, filtering, and pumping equipment, the equipment to service all the ground equipment directly involved in the flight as well as check-out equipment for the airborne equipment, handling equipment of all sorts, and finally the launching pad itself. *See* LAUNCHING PAD COMPLEX. [J.R.SE.]

Flocculation

The formation of larger particles of a solid phase dispersed in a solution by the gathering together of smaller particles. The process whereby minute initial aggregates in a solution develop spontaneously into particles of a new stable phase is known as nucleation. As these particles are growing to visible size, they are known as colloids. Colloidal solutions are frequently stable because the colloid particles do not coalesce to form larger particles by virtue of their similar residual charges, which cause them to repel each other. By changing the ionic environment in which colloid particles exist, for example by adding salts which provide multicharged ions, the colloid particles flocculate. Silver chloride is a good example of a compound which can be precipitated as a flocculated colloid. *See* COLLOID; NUCLEATION; PRECIPITATION (CHEMISTRY); SOLUBILITY PRODUCT CONSTANT. [L.G.]

Flood plains

The valley floors formed by alluviating rivers which are subject to overflow. They may be narrow, between the valley walls of a small stream, or extremely wide, as for example, 125 miles in the lower Mississippi Valley near Memphis. Toward the sea, flood plains may grade imperceptibly into deltas or deltaic coastal plains. The flood plain of the Nile extends 700 miles below the first cataract at Aswan to the Nile delta. Though it is not much wider than 10 miles at any point, the flood plain and delta account for the support of practically all of Egypt's dense population. Ancient potamic civilizations in China, northern India, Mesopotamia, and Egypt arose along the flood plains of the Sian, Indus, Tigris-Euphrates, and Nile, where fertile alluvial soils and dependable water supply favored agriculture. Dikes and drainage canals now prevent or restrict natural flooding, but in protecting cities

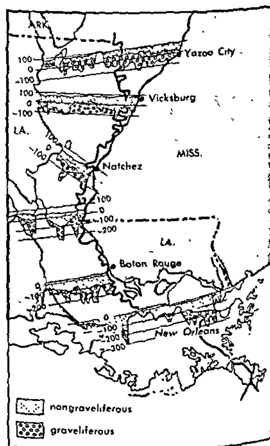


Fig 1. Depth of sediments at various points in Mississippi Alluvial Valley. (*After H. N. Fisk, Mississippi River Commission, U.S. Army, Corps of Engineers, 1944*)



Fig. 2. Oxbow lakes, White River, near Des Arc, Arkansas, 1949. (*USGS*)

and habitations, they rob flood plains of the beneficial alluviating effects of overflow. *See* RIVER ENGINEERING.

Valley floor deposits. The depth of alluvium in wide flood plains toward oceanic coasts commonly amounts to 400 ft or more (Fig. 1), a thickness which is a measure of the volume of water returned to the oceans by melting of continental ice masses during the last 12,000, or possibly 18,000, years. Gravels ordinarily occur abundantly toward the base of the alluvial section and are commonly overlain by finer materials such as sand or silt. Braided streams, with numerous branches around bars and lenticular islands, such as now exist on flood plains where sand and coarser sediments predominate, deposited the coarser materials at depth. Later, when gradients and loads were reduced, many flood-plain rivers adopted sinuous, meandering courses through finer grained alluvium.

Flood-plain surface features. On a section across a typical wide flood plain highest elevations commonly occur at the crests of natural levees which flank the immediate sides of channels. Levee backslopes flatten gradually toward backswamp basins where network drainage, irregular lakes, and high water tables are characteristic and land vegetation may be grassy marsh or tree-covered swamp (*see* PLAINS). Tongues of crevasse deposits, produced during one or more floods, extend natural levee widths irregularly toward the basins. Natural levees are ordinarily a mile or more in width along the lower Mississippi and crests rise above levels of backswamps by 5 ft toward the mouth of the river, over 20 ft at New Orleans, and more than twice that elevation at various points farther upstream. Crevasse deposits extend the useful, solid, fertile, and high land of natural levee belts laterally for 10 miles and more at various places.

Meander-sediment relationships. Meandering rivers continually shift their courses as bars grow around the convex banks of points. These deposits, which accumulate mainly during floods, tend to narrow the channel, causing it to deepen and thus promote undermining of the concave cut banks across from the points. In this way the lower Mississippi each decade abandons about half its channel area to occupy an equal amount of new space, mainly at the expense of cut-bank natural levees. This process also accentuates the size of meander loops, some of which are about 20 miles long. Cut banks, both up- and downstream, however, gradually approach each other, narrowing the neck of the loop, eventually creating a natural cutoff. The abandoned channel becomes a cutoff lake, or oxbow (Fig. 2). These lakes slowly fill with sediment containing considerable clay, to become clay plugs—crescentic masses which erode much more slowly than other types of alluvium (Fig. 3). As they are aligned along both sides of the channel, with horns of the crescents toward the river, they serve effectively to restrict channel migrations to definite meander belts.

It is rare to find diversions from such meander belts along the lower Mississippi and other rivers

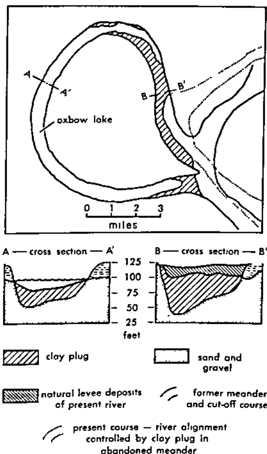


Fig. 3. Clay plugs in abandoned meander. Cross sections along lines A-A' and B-B'. (*After H. N. Fisk, Mississippi River Commission, U.S. Army, Corps of Engineers, 1944*)

flanked by clay plugs. On many other rivers, however, flood plains lack fine materials and are incapable of trapping themselves in meander belts. Such is the case along the Great Meander River of Anatolia where the flood plain is sandy and diversions from the main channel of the river have been numerous.

Other characteristics. Natural levees, definite meander belts, and numerous cutoff lakes are not characteristic of flood plains of rivers that transport little or no bed-load, or which differ little in stage between low water and extreme flood. The smaller rivers of Belgium are examples. Many flood plains are modified by alluvial cones of tributary streams. The Great Meander above Nazilli is forced against its south valley wall by an overwhelming alluvial apron which has been deposited by tributaries on its northern right side and has developed only a narrow flood plain. In a similar manner the San Joaquin has been forced westward by debris from the Sierra Nevada Cones of tributary streams of the lower Rhone account for alternations between steeper gradients, where the Rhone braids, and flatter stretches upstream from each cone, where the Rhone meanders. *See* COASTAL LANDFORMS; DELTA; FLUVIAL EROSION CYCLE; SEA LEVEL FLUCTUATIONS; STREAM TRANSPORT AND DEPOSITION.

[R.J.RU.]

Bibliography: H. N. Fisk, *Geological Investigation of the Alluvial Valley of the Lower Mississippi River*, Mississippi River Commission, U.S. Army, Corps of Engineers, 1944; R J. Russell, Geological geomorphology, *Bull. Geol. Soc. Am.*, 69(1):1–22, 1958.

Floor construction

The selection of the type of construction for the floors of a building should be based on the building's architectural and structural requirements and on cost.

A medium-priced home might contain a wooden floor supported by wooden joists. A floor system frequently used in small commercial buildings, apartment buildings, and higher-priced homes consists of a solid concrete slab supported on steel or concrete beams or open-web steel joists. Open-web joists are available for clear spans up to 120 ft. A thin, lightly reinforced concrete slab cast in one piece with its supporting reinforced concrete joists is particularly economical for short spans. The joists are formed by using metal pans, or clay-tile, gypsum-tile, or concrete-block fillers. A grid system in which the reinforced concrete joists run in two directions may be economical for floor panels that are approximately square.

Light-gage steel cellular floor decking is often used for steel-framed buildings. One type of unit consists of a flat sheet of steel spot-welded to a second sheet bent into a series of troughs. Another consists of two sheets of steel bent into a series of troughs and spot-welded together so that the troughs of one sheet lie against the shoulders of the other.

[C.N.C.]

Floriculture

That area of horticulture concerned with the commercial production and sale of cut flowers, potted flowering and foliage plants, arrangement of flowers, and design, planting, and care of gardens. There is no sharp line of demarcation between the kinds of plants handled by florists and by nurserymen. In temperate regions many florist crops are produced in greenhouses, but in the South and on the West Coast, the same plants may be grown by nurserymen and used in landscape plantings because of the milder climatic conditions.

Florist crop production. Successful production of florist crops is contingent upon the interrelation of such cultural factors, as light, temperature, moisture, and soil fertility. Depending upon the species, propagation is by seed, cuttings, or grafting. Addition of organic matter to and sterilization of the soil are desirable practices for improving plant growth. Periodically, fertilizers are systematically applied to the soil, and the watering of the soil is regulated to provide moisture with adequate aeration.

Flowering of some plants depends on photoperiod (length of day) and on night temperatures. For example, the chrysanthemum, which normally flowers in the fall (a short-day plant), is available in flower every day in the year if the proper photoperiod and temperature treatments are provided. Temperature is also important in governing the time of maturity of nonphotoperiodic plants as well as the quality of all florist crops. Control of pests and diseases is of paramount importance, but conspicuous residues of control materials on the foliage or flowers are particularly objectionable to the retail florist and ultimate customer.

In northerly areas most florist crops are produced in greenhouses enclosed in glass or plastic, but in summer some plants may be grown outside. The more important cut flowers grown outside in summer in the Temperate Zone are dahlia, gladiolus, liatris, and peony; however, in California and Florida, where the climate favors year-round growth of plants, a wide selection of field grown cut flowers is produced.

Flower arrangement. The proper arrangement of flowers is based on principles of design including proportion, balance, and harmony or contrast of color and texture. The commercial flower designer employs these design principles and strives to achieve an attractive yet profitable arrangement, whereas the amateur gardener is concerned only with the esthetic nature of the floral arrangement. *See* FERTILIZING; GRAFTING OF PLANTS; PHOTOPERIODISM IN PLANTS; PLANT DISEASE; PLANT GROWTH; REPRODUCTION, PLANT.

[D.C.K.]

Flotation

A process used to separate particulate solids by causing one group of particles to float while the remaining do not. It utilizes differences in surface-chemical properties of the particles, some of which

are hydrophilic, and are entirely wetted by water, while others are more or less hydrophobic, and are not wetted by water. Although the process is primarily applied to the treatment of minerals, it can also be applied to chemical and biological materials.

Fields of application. Flotation was first used to process sulfide ores, the sulfide minerals being recovered together in a collective concentrate. When emphasis was placed on desirability of collecting different metals in different concentrates, the art progressed to the stage of selective sulfide flotation. This is more delicate, but in view of its great economic advantages it is now practiced almost exclusively. Another application has been to ores of oxygen minerals, such as calcite ($CaCO_3$), barite ($BaSO_4$), hematite (Fe_2O_3), and quartz (SiO_2). Selective oxide flotation is more difficult than selective sulfide flotation. Widespread use is made of flotation for cleaning coals.

A more recent application is to the separation of highly soluble substances such as potassium chloride from sodium chloride. For salines the process requires that a saturated water solution be used. Startlingly good separations have been obtained, and from them a new chemical industry has grown.

Application of flotation to biological separations, to the paper-pulp industry, and to the purification of various wastes are some of the newer uses.

Principles. The finely divided mixed particles to be separated are made into an aqueous suspension. Specific chemicals are added to endow some of the species of solids with a hydrophobic envelope while retaining or emphasizing the hydrophilic character of other solids. The dilution of the suspension is adjusted to the desired point and the mixture is agitated by introduction of air bubbles. The hydrophobic particles and bubbles associate to form "mineralized" bubbles which rise to form a froth laden with selected solids. The froth is skimmed off and conveyed away from the separating vessel while the underflow is likewise conveyed from the separating vessel. New feed is introduced constantly to compensate for the removal of the separated products. The process can be repeated on either the float or the underflow to improve the quality or to complete the separation. In addition, either the underflow or the float, or both, can be treated in a new chemical environment to effect other separations.

In a typical application (for example, to a lead-zinc ore) the mixture can be made to yield first a lead bearing froth. Following a change in chemical conditions whereby the zinc mineral is made hydrophobic, a zinc-bearing froth can be removed. The underflow from the second flotation step may contain iron sulfides together with silicate and carbonate gangue minerals. If the operation is situated in a region in which an iron sulfide concentrate is economically valuable, a third flotation step may be used to float the iron sulfides from other minerals, or even the iron sulfides from each other.

In flotation the following information is important: phase identities; structures and textures of phases within individual particles; particle-size distribution; hydrodynamics of suspensions, aerated suspensions, and froths; hydrophobicity of every solid phase, not only in its natural condition, but also as affected by added chemicals; surface reactions between the particulate solids and the surrounding fluids, both in respect to equilibrium and to kinetics; surface properties of the liquid-gas interface, in respect to equilibrium and to kinetics.

Phase identification and particle structure. In systems studied for their scientific objectives the phases and particle structures are known. On the other hand, in practical cases phases and structures are uncertain and variable. Thus, the sizes of crystals, their shapes and habits, the presence of decomposition products adhering to major phases, and the intergrowth of particular phases, all affect flotation behavior.

The first step in flotation treatment of an ore is to obtain an accurate and comprehensive ultimate chemical analysis. The second step, phase analysis, requires application of the techniques of crystallography, optical mineralogy, and quantitative mineragraphy. Phase analysis should be factual and not based upon assumptions combined with the ultimate chemical analysis. The third step is to learn the texture and structure of the particulate material after reduction to flotation size.

Particle size distribution. Particle size influences *hydrodynamic behavior, probability of attachment* to bubbles, as well as engineering factors such as machine capacity, selectivity, degree of phase liberation, and facility of product dewatering.

Particles are not all the same size. In fact they may range from about 10^{-1} to 10^{-5} cm, or even more widely. With ore pulps, particle size distribution is a consequence of comminution and of the peculiarities of structure and texture of the ore. With materials other than ore pulps, particle size distribution again reflects the past history of the material treated.

Hydrodynamics. Movement of particles and bubbles in still water depends primarily on their size, shape, and specific gravity, also on temperature and on the boundaries of the container, its size compared to that of the particles, and the crowding of the fluid by the particles. In agitated water, the

and position.

Mechanisms for the formation of gas-solid aggregates are several, including encounter of bubbles and particles either in the pulp body or the bubble column, and gas precipitation. These phenomena involve nucleation processes about which little is known. Contact-angle magnitude, size, shape, and speed of bubbles and particles are among the principal factors influencing the formation of gas-solid aggregates.

Hydrophobicity. Attachment of gas to a solid surface under water may not occur, no matter how much a bubble is pressed against the solid nor for how long. This is in the case of complete wetting of the solid by water. On the other hand, attachment may occur even in a short fraction of a second (10^{-3}–10^{-6} sec), with formation of a contact angle. The contact angle, which is measured normally to the three-phase line of contact of the air, water, and surface of the particle, relates the three surface energies by the Young equation:

$$E_{SG} = E_{SL} + E_{GL} \cos \theta$$

In this relation, the terms in E are specific energies, G, L, S refer to the gas, liquid, and solid respectively, and θ is the contact angle.

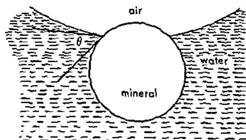


Fig 1. Suspension of a spherical mineral particle at the surface of a large bubble.

Complete hydrophobicity would be obtained for $\theta = 180°$. No such case is known, one of the larger values for θ (paraffin) being about 104°. A contact angle of 90° (which denotes indifference of the solid toward the two fluids, air and water) nevertheless permits very high floatability. In fact a floatability that is not nil ...

... through absorption at their surface of some ...

... cules, fo ... gree of ... fluorite (CaF_2) or an endless array of covalently bonded atoms such as is found in quartz has no ...

... that are capable of forming a hydrocarbon-bearing ion. The hydrocarbon-bearing ion, or the acid or base derived from it, adsorb at the surface of the affected solid, forming on it a partly complete monolayer orientated with the hydrocarbon groups away from the solid.

Examples of collectors are soaps, xanthates, and amines. Soaps give hydrocarbon-bearing anions readily fixed by surfaces containing alkaline-earth or heavy-metal ions. Xanthates give hydrocarbon-bearing anions readily fixed by surfaces containing metals that form very insoluble sulfides. Amines give hydrocarbon-bearing cations fixed more or less readily by surfaces that contain anions capable of forming insoluble amine salts or by surfaces containing ions sterically favorable for exchange with the particular aminium ion involved.

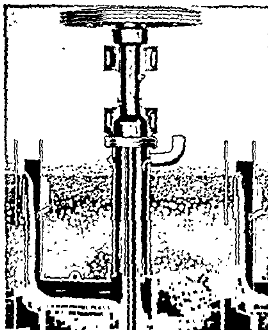
Induced floatability is typically obtained under conditions that lead to distribution of the collecting agent between liquid and interface such that the adsorbed layer is only a partly complete monolayer. For example, barite ($BaSO_4$) may be floated with a soap (sodium laurate) provided the reagent concentration, barium-ion content in solution, and pH ... laurate monolayer ... of the value needed to justify the appearance of a barium laurate phase.

Selection by flotation agents. Controlled addition of collector may be insufficient to produce the desired spread in properties between the various solid phases. Added reagents which have proved useful in aiding selection include salts of heavy metals (for example, to introduce Cu^{++}), sequestering agents (such as NaCN to remove Cu^{++}), pH-controlling agents, and protective colloids. Success of more than one operation depends on various modifiers added in proper quantity and place.

Surface properties of liquid-gas interface. Frothing agents, added to form fine bubbles, are heteropolar organic substances with a hydrocarbon group at one end of the molecule and a nonhydrocarbon group at the other. Frothers differ from collectors by their nonionizable character. A moderately long-chained alcohol, ketone, or ether makes a good frother. More generally such an agent contains a properly sized hydrocarbon group to give the molecules the desired balance between the hydrophilic character of the polar group and the hydrophobic character of the hydrocarbon group.

Frothers are adsorbed strongly at the boundary ... ation is time- ... ich ... and gas and across the interface), the boundary is endowed with an elasticity which greatly reduces bubble-bubble coalescence. This makes froths possible, thereby increasing the practical capacity of flotation machines and facilitating practical operation.

Flotation can be used as a continuous process, and it can be controlled by inspection and analysis of the froth and of the final underflow. Reagent addition is sometimes automatically controlled by some feature of the pulp; for example, if it is decided by prior experimental work that the operation requires maintenance of a pH between 9.5 and 10, an automatic pH-recording device can be used to mastermind the feeder for lime or soda ash. Automatic continuous chemical analysis, with feed-



F . . . . of a flotation cell. (Denver Equipment Co.)

back control of operating variables, should occur in the near future. The quantity of material treated by flotation annually exceeds 200,000,000 tons. See ADSORPTION; INTERFACE OF PHASES; ORE DRESSING; SIZE REDUCTION; SURFACE-ACTIVE AGENT; UNIT OPERATIONS. [A.M.C.]

Bibliography: A. M. Gaudin, *Flotation*, 2d ed., 1957; K. L. Sutherland and I. W. Wark, *Principles of Flotation*, 1955.

Flounder

Any of a number of marine fishes belonging to the family Pleuronectidae, and readily recognized by their asymmetrical shape and flattened body. Flounders begin life swimming upright and are symmetrical. When a few weeks old, the young fish turns over on one side and the cranium becomes twisted, placing both eyes on the same side, which then develops into the dorsal surface of the animal. The long dorsal and anal fins thus become continuous



T. . . er flounder, *Pseudopleuronectes americanus*; length to 18 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

with the flattened outline of the fish. Flounders are all bottom dwellers and usually live in shallow, coastal, sometimes brackish, water. They are excellent food fish and provide some sport fishing. Soles, plaice, flukes, halibuts, and turbots are all flounders. Various species are of considerable value in the commercial fishery. See PLEURONECTIFORMES.

[J.D.B.]

Flow measurement

The determination of the quantity of a fluid, either liquid, vapor, or gas, that passes a given point in a pipe, duct, or open channel. Flow may be expressed in units of rate, such as ft^3/sec, gal/min, lb/hour, or in terms of total quantity (integrated rate of flow) such as cubic feet, gallons, or pounds.

Flow measurement, though centuries old, has only been developed as a science since the advent of the industrial age. The development was due to the need for controlled process flows, stricter accounting methods, and more efficient operations, and to the realization that control of flow is a simple and convenient method to control other process variables, such as temperature and pressure.

Measurement is accomplished by a variety of means depending upon the . . . end . . . e . . . is . . . ta . . . generates a signal, while a secondary device translates this signal into a motion or a secondary signal for indicating, recording, controlling, or totalizing the flow. Other devices indicate or totalize the flow directly through the action of the flowing fluid on a float or on a rotating impeller or other device placed in the fluid stream. The table shows several types of metering devices and the principles upon which they operate.

RATE-OF-FLOW VOLUME METERS

These met . . . the actua . . . of a phen . . . with the flow. In head meters, flow through the primary device creates a differential pressure; in velocity meters the rotation of an impeller or rotor is caused by the flow.

Variable-head meters. By far the most common in general use, these are based on the law of conservation of energy, which requires that the total energy at any given point in a stream is equal to the total energy at a second point in the stream, neglecting losses between the points. While it is possible to convert pressure (potential energy) to velocity (kinetic energy) and vice versa, the total amount of energy does not change. By use of a restriction in the pipe, such as an orifice plate (Fig. 1), or a Venturi tube (Fig. 2), a portion of the potential energy of the stream is temporarily converted to kinetic energy as the fluid . . . pass . . . . . . ated to flow by the sim-

Examples of fluid meters

Name or type	Operating principle	Useful flow range	System accuracy
Rate of flow volume meters			
Orifice (Fig. 1 and 4) Venturi tube (Fig. 2 and 5) Flow nozzle (Fig. 6)	Constant area and variable pressure drop	3½:1	±1-1½%
Pitot tube (Fig 7) Pipe elbow (Fig 8)	Constant area and variable pressure drop	3½:1	±5%, unless flow calibrated
V-notch rectangular weirs (Fig. 13) Venturi or Parshall flume (Fig. 13)	Variable head and variable area	10:1	±3%
Rotameter (Fig. 14) Piston (Fig. 15)	Variable area and constant pressure drop	10:1	±1-2%
Electromagnetic meter (Fig. 16)	Electromagnetic induction	20:1	±1%
Acoustic velocity	Difference of sound velocities going upstream and down	20:1	±5%
Salt velocity	Measurement of transit time along a fixed distance	10:1	±1-1½%
Quantity meters			
Nutating disk (Fig. 17) Rotating vane (Fig. 17) Gear or impeller (Fig. 17)	Positive displacement, separation into and counting of discrete fluid quantities	20:1	±0.5%
Propeller (Fig 18) Turbine (Fig. 18) Vortex cage (Fig. 18)	Transformation of linear velocity to rotation	15:1	±0.5%
Rate of flow mass meters			
Axial flow (Fig 19)	Momentum effect due to change in direction of fluid flow path	10:1	±1%
Gyroscopic type (Fig 20)	Momentum effect due to change in direction of fluid flow path	10:1	±1%
Density-compensated volume flow meters	Volume flow meter plus densitometer	3½:1	±2%

plified hydraulic equation: $V = K\sqrt{2g\ \Delta H}$, where V is the velocity of flow through the restriction, K an experimentally determined proportionality constant (flow coefficient) relating the actual velocity to the theoretical velocity expected, and g, the acceleration of gravity. Volume flow rate Q equals AV, where A equals the cross-sectional area of the restriction. The mass (or weight) flow rate, $W = \gamma AV$, where γ is the fluid density in lb/ft³. This gives $W = 359\ K\ d^2\ \sqrt{h_w \gamma}$ in common engineering units, where W is the flow rate in lb/hour, d is the orifice diameter in inches, and h_w is the differential pressure in inches of water. Values of K are given in texts listed in the bibliography.

Beyond the restriction, the flow particles slow down until the velocity is once more the same as that in the conduit upstream of the restriction, except for the effect of turbulent and frictional losses associated with the acceleration and deceleration at the restriction. The magnitude of the loss depends upon the configuration of the primary device (Fig. 3). *See* ORIFICE.

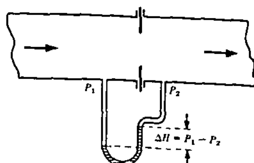


Fig 1. Orifice meter.

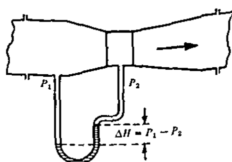


Fig 2. Venturi meter.

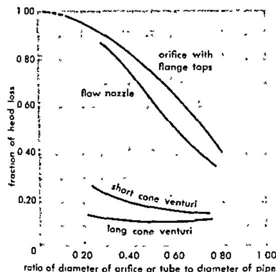


Fig 3. Relation of pressure loss to diameter ratio for common variable head meters.

Orifice plate. The common orifice plate is a thin plate inserted between pipe flanges, usually having a round, concentric hole with a sharp, square upstream edge. The extensive empirical data available make it the most common of the primary devices. Segmental and eccentric orifices are useful for the measurement of liquids with solids, and of vapors or gases with entrained liquids. Orifice plates are shown in Fig. 4.

Venturi tube. This device, shown in Fig 5, has the advantage of introducing less permanent loss of pressure than the orifice plate (see Fig. 3). The converging inlet cone permits solids and dirt to be flushed through it, and the outlet cone reduces the turbulent losses. *See* VENTURI TUBE.

Flow nozzle. Figure 6 shows that the flow nozzle also has a converging inlet, but it has no diverging outlet section. The pressure loss is similar to that for an orifice plate. It is used where solids are entrained in the liquid and also where fluid velocities are particularly high (it will pass about 60% more fluid than an orifice of the same diameter).

Pitot tubes. In industrial flow measurement, pitot tubes are limited to use in spot checks and for comparative measurements to determine trends. This device (Fig. 7), consists of a piece of tubing, inserted into the pipe or channel, with a hole in the end directed upstream. The impact pressure developed by the flowing stream striking the hole is conveyed outside to the secondary device where it is converted to an indication of fluid velocity at the location of the tip. Several readings should be taken along a diameter (preferably two diameters) of the pipe and the average velocity determined. The pitot tube is used extensively as a laboratory device for spot velocity measurement and also for speed measurement of vehicles and in wind tunnel work. *See* AIR-VELOCITY MEASUREMENT; PITOT TUBE.

Pipe elbow. Fluid flow around the bend in a pipe elbow (Fig. 8) produces a centrifugal force greater

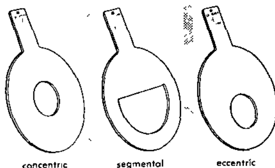


Fig. 4. Orifice plates

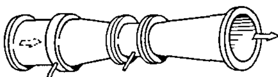


Fig 5. Venturi tube.

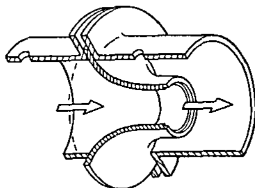


Fig 6. Flow nozzle

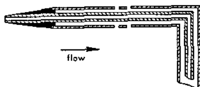


Fig. 7. Laboratory type pitot tube.

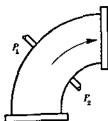


Fig 8. Pipe elbow.

on the outside than on the inside, and thereby produces a difference in pressure between pressure connections on the inside and outside of the bend. This differential pressure is used as indication of the volume flow rate. This device is often used for rough flow measurement because it is not necessary to disturb existing piping.

Differential pressure. All these primary devices create a difference in pressure between two points. This differential pressure must be measured and converted by the equations given to obtain flow. Differential pressure may be measured accurately by simple liquid-filled (usually mercury) U-tube manometers (*see* MANOMETER), or by more refined types of meters, such as the float-operated mechanism or the weight-balanced, ring-type meter, to provide controlling, recording, and totalizing functions. These devices (Fig. 9) are ordinarily connected directly to the pressure connections on the primary device and are therefore exposed to the process fluids. Sometimes an inert liquid seal fluid is used to isolate the secondary device from corrosive process fluids. In either case, the manometers are best located adjacent to the primary device These manometers are accurate if the mercury is clean and the fluid flow rate is not changing too rapidly.

Diaphragm meters. Diaphragm and bellows-type differential pressure measuring devices were developed to eliminate manometer fluids, and to provide faster response and easier installations. One type (force balance), Fig. 10, causes the variation of an air pressure in a small bellows to oppose the net force created by the differential pressure on opposite sides of a diaphragm or diaphragm capsule. This air pressure is used as a signal representing the flow rate.

Another type (deflection), Fig. 11, places the differential pressure on opposite sides of a bellows or diaphragm capsule and, allowing the bellows or diaphragm to move against a calibrated spring, produces a motion directly proportional to the differential pressure. This motion is used directly or is

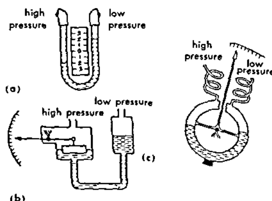


Fig. 9. Differential-pressure meters. (a) U-tube manometer. (b) Mercury float meter. (c) Weight-balanced ring type meter.

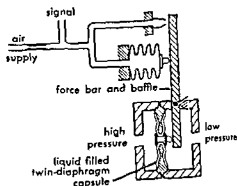


Fig. 10. Diaphragm meter, force balance type.

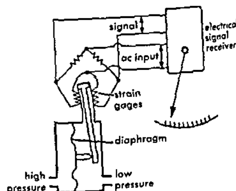


Fig. 11. Diaphragm meter, deflection type.

measured by strain gages or an inductive pickup

Both force balance and deflection meters are well suited to either pneumatic or electrical transmission of the flow signal.

Since these secondary devices measure the differential pressure developed, they must be fitted with square-root scales and charts based on the equations given in order to read flow. These scales are particularly difficult to read at low values. Uniform flow-rate scales are obtained with manometers with specially shaped chambers or with shaped plugs (Ledoux bell), as shown in Fig. 12, or with special cams in the linkage between the manometer and the indicating pointers or pens.

Variable-head, variable-area meters. Flow in open channels is obtained by measuring the height of liquid passing through a weir or flume (*Fig.* 13). Consistent results require that the flowing stream approach the measuring device with a negligible velocity. Because the height-flow velocity relationship is exponential, the secondary device frequently uses cams to obtain a uniform flow scale.

Variable-area, constant-head meters. These fall into two general classes. In the rotameter (Fig 14), the fluid flows upward through a tapered tube, lifting a shaped weight (possibly misnamed a *float*) to a position where the upward fluid force just balances its weight, and thereby indicates the flow rate.

In the piston type (Fig. 15), the buoyant force of the liquid carries the piston upward until sufficient area has been uncovered in a slot in the side

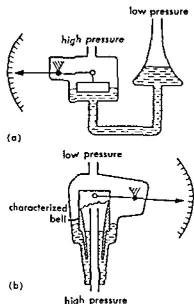


Fig 12. Meters with uniform flow rate scales. (a) Shaped-chamber manometer. (b) Ledoux bell.

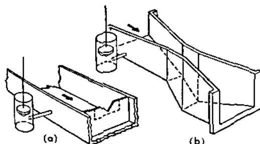


Fig 13. Variable-head, variable-area meters. (a) Weir. (b) Flume.

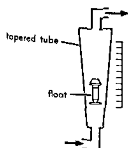


Fig. 14. Rotameter-type, variable-area, constant-head meter.

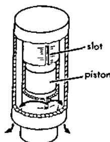


Fig 15 Piston-type, variable-area, constant-head meter.

of the vertical tube to allow the liquid to flow through the slot. The position of the piston indicates the flow rate and, by properly shaping the slot, this type meter may be made with a uniform flow scale. Mechanical, pneumatic, and electrical position-sensing mechanisms are available to make recorders, indicators, and totalizers of these devices.

Electromagnetic meter. A magnetic field is applied across the pipe carrying a flowing liquid, and the voltage generated between two perpendicularly located electrodes contacting the liquid is measured as in Fig. 16. Since this voltage is strictly

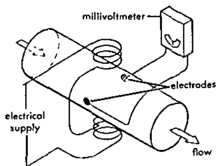


Fig. 16. Electromagnetic meter.

proportional to the velocity, a linear flow record or indication results. Practical flow measurements can be made on any liquid that is at all conductive. The linear scale allows accurate flow measurements over a greater range than is possible in the differential pressure meters. The lack of obstruction to fluid flow makes it possible to measure thick slurries and gummy liquids

Sonic velocity flow meter. The velocity of sound in a fluid is different in the direction of flow and in the direction counter to flow; this difference is indicative of the flow rate. This meter can be used in large channels. The electronic circuitry required to correct for ambient effects makes this type of flow measurement too expensive for normal installations.

Salt velocity method. In this classic method a small quantity of salt or radioactive isotope is injected into the flowing stream, and the transit time of passage between two fixed points is found by measuring electrical conductivity or radiation level at those points. Inaccuracies are introduced by variations in flow cross section in the pipe or channel being measured. The method, however, is an easy one to use for a quick check under comparatively adverse conditions.

QUANTITY METERS

Positive displacement meters. The flowing

normally used to indicate the integrated flow

through the meter. Figure 17 shows three types of positive displacement meters.

Propellor, turbine, and vortex cage meters. These meters exert only slight retardation on the flowing fluid. The elements rotate at a speed that is linear with fluid velocity. Revolutions are counted through a mechanical or magnetic coupling, the rate of rotation being volume flow rate (Fig. 18).

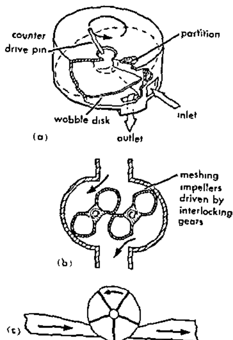


Fig 17. Positive displacement meters. (a) Nutating disk. (b) Lobed *impeller* (c) *Rotary* vane.

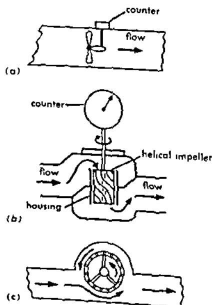


Fig 18. Quantity meters. (a) Propeller, (b) *Turbine*, (c) Vortex.

Flow correction factors. The accuracy of calibration of the above meters is dependent, to varying degrees, on the temperature, pressure, and the process fluid physical characteristics. Accurate flow coefficients include terms for viscosity and density, and the readings of these meters must be corrected for any changes if accurate measurement is required. It is often the practice to take pressure and temperature measurements of vapors and gases, or temperature measurement of liquids, at the flow metering device so that corrections may be applied either by the mechanism of a recording or totalizing device or introduced manually. Some meters are available in which these corrections are introduced automatically with great accuracy while others are made in which only a first-order approximation is introduced. Proper selection is determined by the requirements of the installation. A meter compensated for fluid conditions can become a mass flow meter

RATE-OF-FLOW MASS METERS

Three types. Several mass flow meters measure the actual mass flowing and require no compensation. Most of them operate on the principle of conservation of angular momentum. The torque required to impart or to remove a constant angular momentum to the fluid is directly proportional to the mass rate of flow. In one type, the fluid in the pipe is made to rotate at a constant speed by a motor-driven impeller. The torque required by a second, stationary impeller to straighten the flow again is a direct measurement of the mass flow (see Fig. 19). In another configuration (using the

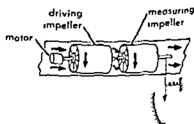


Fig. 19. Axial flow type, mass meter.

Coriolis effect) a ribbed disk is fastened by a torque-sensing member within a housing and both are rotated at constant speed. The fluid enters at the center of the disk and is accelerated radially; the torque on the disk is a direct measure of the mass flow rate. Unfortunately, the practical problem of maintaining the rotating seals in the device (as well as in the gyroscopic type) has hindered their commercial development.

A third type is based on the principle of the gyroscope (Fig. 20). The flowing stream is sent through a pipe of suitable shape so that the mass of material flowing corresponds to a gyro wheel rotating about axis C. The entire gyrolike assembly is rotating about axis A and a torque is produced about axis B proportional to the angular momentum

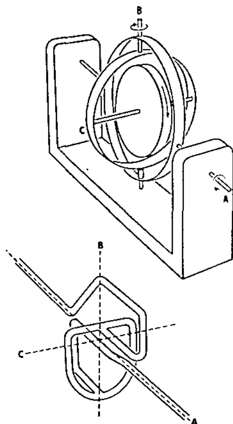


Fig. 20. Gyroscope type meter.

of the simulated gyro wheel. This torque is therefore directly proportional to the mass rate of flow. The pretzel-like configuration of pipe is introduced to eliminate centrifugal and other extraneous effects.

Accuracy of measurement. The accuracy is generally taken of the total system of primary and secondary devices and customarily is expressed in percentage of full-scale value. Accuracy of flow measurement is poor at the low end of the square-root scale of differential pressure meters. Rangeability of a flow meter refers to the ratio of full-scale flow to a practical minimum flow rate. A figure of 3½-1 is normally used for square-root scale meters, while linear meters, such as the magnetic flow meter, normally have figures of 20:1 or higher. Meters with rotating parts in the flowing stream, although linear, have a lower useful range because of the effect of friction at low flow rates. See INSTRUMENTATION; PHYSICAL MEASUREMENTS

[L.P E.; M.BR.]

Bibliography: American Society of Mechanical Engineers, *Fluid Meters—Their Theory and Application*, 5th ed., 1959; D. M. Considine (ed.), *Process Instruments and Controls Handbook*, 1957; L. S. Marks, *Mechanical Engineers' Handbook*, rev. ed., 1958; L. K. Spink, *Principles and Practice of Flow Meter Engineering*, 8th ed., 1958.

Flow of fluids

Motion of a fluid as a continuum. Fluids flow, whereas solids move as bodies. In flow, the individual particles of a substance move relative to each other as well as to their surroundings. A fluid is a substance that flows under the slightest stress. Thus, its own pressure is sufficient to cause a gas to flow throughout a container to which it is admitted. The weight of a liquid is sufficient to cause it to flow in a container but without significant change in volume Considerable external force is required to cause a solid to flow.

By analogy, electricity, heat, and other forms of energy are said to flow (*see* ELECTRICAL CONDUCTION IN GASES, HEAT TRANSFER). This article discusses only flow of fluids.

Fluids. A fluid may be liquid, vapor, or gas. A liquid will fill the container which holds it, but it may have a free surface, that is, a surface from which all pressure is removed except that of its own vapor. All liquids are relatively incompressible

A vapor is a gas whose temperature and pressure are such that it is very near the liquid phase. Thus, steam is considered to be a vapor because its state is not far from that of water.

A gas may be defined as a highly superheated vapor; that is, its state is far removed from the liquid phase. Thus, air is considered to be a gas because its state is normally very far from that of liquid air.

A gas is very compressible, and when all external pressure is removed, it tends to expand indefinitely. A gas is therefore in equilibrium only when it is completely enclosed.

The volume of a liquid is altered only slightly by changes in either pressure or temperature unless the temperature change is considerable or unless the initial temperature is near the critical temperature, but the volume of a gas or a vapor is greatly affected by changes in either pressure or temperature. *See* GAS; LIQUID.

Ideal and real flow. When there is no friction between adjacent moving particles, flow is termed ideal; that is, viscosity is zero In ideal flow, internal forces at any section are always normal to the section Forces are purely pressure forces. Such flow is approached but never achieved in reality.

In a real fluid, tangential or shearing forces always come into being whenever motion takes place, thus giving rise to fluid friction, because these forces oppose the sliding of one particle past another. These friction forces are due to a property called viscosity.

Viscosity. The viscosity of a fluid is a measure of its resistance to shear or to angular deformation. Consider two parallel plates large enough that edge conditions may be neglected and placed a small

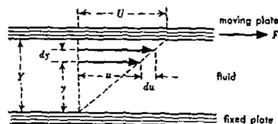


Fig 1 Viscous flow.

moved relative to it with a velocity U by the application of a force F corresponding to some area A of the moving plate. Particles of the fluid in contact with each plate will adhere to it, and if the distance Y is not too great and the velocity U not too high, the velocity gradient will be a straight line. The action of the fluid may be likened to a series of thin sheets, each of which slips a little relative to the next. Experiment has shown that for a large class of fluids the shear stress τ between any two thin sheets of fluid may be expressed as

$$\tau = \frac{F}{A} = \mu \frac{U}{Y} = \mu \frac{du}{dy}$$

where μ is the coefficient of viscosity (*see* NEWTONIAN FLUID). This is also called the absolute or dynamic viscosity.

In a Newtonian fluid, the viscosity does not *change with the* rate of *deformation* and is represented by the straight line (Fig. 2). The slope of this line is determined by the [illegible] cosity. [illegible] er's ink [illegible] the rate of deformation, but these are usually not of engineering importance.

Absolute viscosity. In the metric system the unit of absolute viscosity *is the poise* (1 poise = 1 dyne-sec/cm²). Because most fluids have low viscosities, the centipoise (0.01 poise) is frequently a more convenient unit. It has the further advantage that the viscosity of water at 68.4°F (1 centipoise) provides a convenient reference.

In the English gravitational, or engineers', system, the viscosity unit in pounds (force) is 1 lb-sec/ft² or 1 slug/ft-sec. In the English absolute system, viscosity is expressed in terms of poundal-sec/ft² or lb/ft-sec. There are no names for the English units.

Conversion factors are 1 poise ≡ 100 centipoises ≡ 0.00209 lb-sec/ft² ≡ 0.0672 poundal-sec/ft².

The viscosities of all liquids decrease with an increase in temperature, whereas those of all gases and vapors increase. The absolute viscosity of both liquids and gases is practically independent of pressure except for extremely high pressures at which there is a small increase.

Kinematic viscosity. In many fluid-flow problems there occurs the value of viscosity divided by density. This is called kinematic viscosity, and is $\nu = \mu/\rho$. In the metric system the unit of kinematic viscosity is the stoke (1 stoke = 1 poise divided by density, g/cm³). A convenient unit is the centistoke (0.01 stoke). The dimensions of kinematic viscosity are cm²/sec.

There is no name for the unit of kinematic viscosity in the English system, but the dimensions are ft²/sec. It is necessary to be consistent in the units [illegible] weight.

Conversion factors are 1 stoke ≡ 100 centistokes ≡ 0.001076 ft²/sec and 1 ft²/sec ≡ 929 cm²/sec. *See* VISCOSITY OF GASES; VISCOSITY OF LIQUIDS.

Density and specific weight. Density is mass per unit volume and is normally indicated by ρ. In the metric system the dimensions of ρ are g/cm³. In the English systems they are either slugs/ft³ or lb/ft³ according to whether the engineers' or the absolute system is used. The value of density is the same for any location.

Specific weight is weight per unit volume and is commonly designated by either w or γ. In the English engineers' system dimensions are lb/ft³. In the absolute system they are slugs/ft³ where 1 slug ≡ 32.174 lb (mass). Specific weight in the engineers' system varies with the value of gravity and hence varies with location.

The relation between density and specific weight is $\rho = w/g$ where g is the acceleration of gravity. The standard sea-level values of g are 980.66 cm/sec² or 32.174 ft/sec². *See* DENSITY; *see also* SPECIFIC GRAVITY.

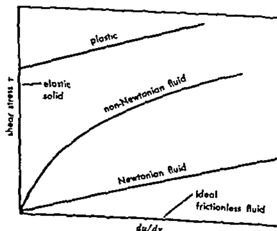


Fig. 2 Types of frictional flow.

Compressibility. All liquids are relatively incompressible, and for most purposes water in particular can be treated as incompressible; *yet it* is ten times as compressible as steel. The passage through water of a sound wave, which is really a

pressure wave, is a result of the water's compressibility. *See* WATER HAMMER.

In dealing with the flow of air or other gases when the change in pressure is small, so that the change in density is negligible, even gases may be treated as incompressible. For an airplane flying at speeds of less than 250 mph the air may be considered to be of constant density (*see* SUBSONIC FLIGHT). However, as the speed approaches that of sound in air, which is of the order of 700 mph, the pressure of the air adjacent to the body becomes materially different from that at some distance away, and air must then be considered to be compressible. *See* COMPRESSIBLE FLOW.

The change in volume from v to v_1 of a fluid as the result of a change in pressure Δp can be determined by $\Delta v/v_1 = -\Delta p/E_v$ where E_v is the mean value of the volume modulus of elasticity for the pressure range. For water the value of the volume modulus varies with both temperature and pressure, but a typical value for most conditions is 300,000 psi (Δp and E_v must both be in the same units). For isothermal compression of a perfect gas, $E_v = p$ and for an adiabatic compression $E_v = kp$ where k is the ratio of the specific heat at constant pressure to that at constant volume. For air the value of k is normally about 1.4, and p is the average pressure range. For low pressures air is many thousand times as compressible as water.

For practical use the change in volume of a gas is better determined by means of the perfect gas equation of state.

Perfect gas. A perfect gas is one whose equation of state is $pV = WRT$ or $pv = RT$, where p is absolute pressure, V is total volume, v is volume per unit weight, W is total weight, T is absolute temperature, and R is a gas constant. A consistent system of units must be used. Thus p may be in lb/ft², V in ft³, v in ft³/lb, W in lb, and T in °F + 459.7°. The value of R in this system is 1546/molecular weight.

There is no perfect gas, but air and other so-called permanent gases (gases at conditions far from their liquid states) may usually be so considered, and for such real gases the product of R times molecular weight may range from 1485 to 1786. For air the value of R is 53.3. Even water vapor in the air follows the above law closely with $R = 1541/18 = 85.6$. For normal steam pressures, this does not hold, and vapor tables must be used instead (*see* STEAM). Likewise, for real gases when high pressure or low temperatures are involved the equation of state is more complicated. *See* VIRIAL EQUATION.

Real gases. For real gases, in which the conditions are such that the perfect gas law is not sufficiently accurate, a compressibility factor Z may be used in the equation of state $pv = ZRT$ where Z may range from 0.2 to 3 in extreme cases.

The van der Waals equation of state is sometimes useful at high pressures. *See* VAN DER WAALS EQUATION.

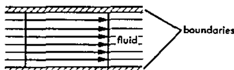


Fig. 3. Ideal fluid flow.

Ideal fluid flow. The flow of an ideal fluid is also termed inviscid flow, because it assumes an imaginary fluid of zero viscosity which therefore has no fluid friction. In such flow all particles of the fluid would move along individual streamlines and with equal velocities (Fig. 3).

Because there is no fluid friction, the total energy is the same throughout. For an incompressible liquid Bernoulli's equation can be used,

$$\frac{p}{w} + z + \frac{V^2}{2g} = \text{constant}$$

from one streamline to the next and also along any streamline. In this equation p/w is pressure head in height of the fluid. If p is pressure in lb/ft² and w is specific weight in lb/ft³, p/w is a linear quan-

linear quantity with V the velocity in ft/sec and g the acceleration of gravity in ft/sec².

Each term in Bernoulli's equation is a linear quantity that also represents energy per unit weight. Strictly speaking, the energy possessed by a particle of fluid is the sum of its potential and kinetic energy; that is, $z + V^2/2g$, whereas p/w represents the work done as a result of pressure and motion, but work and energy are measured in the same units.

In many real cases fluid friction is so small that results obtained by Bernoulli's theorem are sufficiently accurate for practical purposes or in many cases need only be modified slightly. Thus, the velocity of a jet issuing from a tank under a head h may be obtained by introducing a velocity coefficient into the equation such that $V = C_v\sqrt{2gh}$, where C_v may have a value as high as 0.98 or 0.99. Thus, the actual value of V is little less than the

friction is large, and the actual result may be very different from the ideal frictionless value.

The effect of viscous friction really originates where the fluid is in contact with a solid surface, and at a considerable distance from such surface the effect of friction is much less (*see* BOUNDARY-LAYER FLOW). Thus, fluid friction resulting from viscosity produces significant results near a body, such as an airplane, but at some distance away the relative velocity of the air may be considered as uniform (Fig. 3).

Real fluid flow. Real fluid flow and viscous flow are synonymous terms; all real fluids are more or less viscous (*see* VISCOUS FLOW). Because of fluid friction, Bernoulli's equation must be modified by the introduction of a term h_f which represents the loss of head or energy. Thus, an energy equation between two points is

$$\frac{p_1}{w} + z_1 + \frac{V_1^2}{2g} = \frac{p_2}{w} + z_2 + \frac{V_2^2}{2g} + h_f$$

which shows that the total head always decreases in the direction of flow. This equation applies to an incompressible fluid. *See* BERNOULLI'S THEOREM.

In some real cases the loss of head may be small and may be disregarded with slight error, but in other cases it may be large. Values of h_f are obtained for specific cases by equations that are more or less empirical and are based upon experimental data.

Laminar flow. Laminar flow may also be called streamline flow, because all particles of the fluid move in distinct and separate lines (*see* STREAMLINE FLOW). It is called laminar flow because the action is as if layers or lamina of fluid slide relative to each other. In the case of a laminar flow in a circular pipe, the velocity adjacent to the wall is zero and increases to a maximum in the center of the pipe (Fig. 4). The velocity profile in a circular

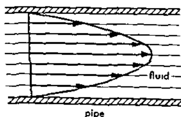


Fig. 4. Laminar flow in a circular pipe.

pipe is a parabola and the average velocity is 0.5 times the maximum velocity in the center. *See* PIPE FLOW.

In the case of laminar flow in a circular pipe the loss of head due to fluid friction is not given by any empirical equation but is given by a rational equation known as the Hagen-Poiseuille law, which is

$$h_f = \frac{p_1 - p_2}{w} = 32\frac{\mu}{w}\frac{L}{D^2}V = 32\nu\frac{L}{gD^2}V$$

where h_f is linear head in feet of the fluid, p is pressure in lb/ft², μ is absolute viscosity in lb/ft-sec, w is specific weight in lb/ft³, ν is kinematic viscosity in ft²/sec, L is distance between pressure levels (1) and (2) in ft, D is diameter in ft, and V is velocity in ft/sec.

Laminar flow in a circular pipe will be found when the Reynolds number, $DV\rho/\mu = DV/\nu$, is less than 2000. *See* LAMINAR FLOW.

Turbulent flow. In turbulent flow no distinct streamlines are found. Instead the fluid consists of a mass of eddies (Fig. 5). No two particles may follow the same or similar paths. The velocity profile

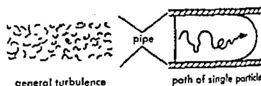


Fig. 5. Turbulent flow in circular pipe.

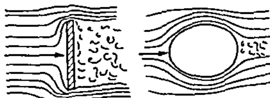


Fig. 6. Turbulent wake.

shows a maximum velocity in the center while near the wall the velocity is about one-half the center velocity. The profile is flatter for a smooth pipe than it is for a rough one, and the ratio of average velocity to maximum velocity in the center ranges from about 0.74 for a very rough pipe to about 0.88 for a very smooth one.

In a circular pipe the loss of head due to fluid friction is given by $h_f = f(L/D)V^2/2g$, where f is a function of both Reynolds number and the wall roughness, L/D is ratio of length to diameter, V is the average velocity in ft/sec, and g is the acceleration of gravity, normally taken as 32.2 ft/sec².

Turbulent flow occurs in a circular pipe when Reynolds number has values greater than 2000.

In the case of a solid body immersed in a stream, there is a turbulent wake in the rear (Fig. 6). This produces a pressure difference which results in a force on the body known as form or pressure drag. *See* TURBULENT FLOW; WAKE FLOW.

Uniform flow. If at a given instant the velocity is the same in both magnitude and direction at every point in space, the flow is uniform. This strict definition can have little meaning for the flow of a real fluid when the velocity varies across a section. However, when the size and shape of cross section are the same in any given length, the flow is said to be uniform. Specifically, the flow in a pipe of constant diameter is uniform; the flow in a pipe of varying size is not. Also, the flow in an open canal is uniform if the size and shape of the cross section of the stream are the same at different locations along the canal. *See* CHANNEL, OPEN; STREAM GAGING.

Steady flow. By steady flow is meant that all conditions at any one point are constant with respect to time. True steady flow is found only with laminar flow. In turbulent flow there are continual fluctuations in velocity and pressure at every point. However, if the values fluctuate on both sides of an average value that is constant, then the flow is mean steady flow.

In steady flow, conditions are usually constant in time from one section to another, although not necessarily the same at different sections. Thus, along

the line of flow the equation of continuity applies, which is

$$W = w_1 A_1 V_1 = w_2 A_2 V_2 = \text{constant}$$

or for a fluid of constant specific weight

$$Q = A_1 V_1 = A_2 V_2 = \text{constant}$$

where W is flow in lb/sec, w is specific weight in lb/ft^3, A is cross-sectional area in ft^2 normal to the velocity, and V is average velocity across the section in ft/sec.

Unsteady flow. This means that conditions are changing with respect to time and ultimately may become either steady flow or zero flow. This changing rate of flow may take place slowly, as when the action of a valve in a channel produces a gradual change in the rate, or it may take place rapidly as a result of a sudden closure, which produces a phenomenon known as water hammer. It is also found in such a case as the flow from one reservoir to another, in which equilibrium is approached as the two levels approach each other.

Unsteady flow also includes periodic motion such as that of waves on beaches, tidal motion in estuaries, and other oscillations. The difference between such cases and mean steady flow in turbulent flow is that the deviations from the mean are much greater and the time scale is also much longer. *See* ATMOSPHERE; ESTUARINE OCEANOGRAPHY; WAVE MOTION IN LIQUIDS; WIND STRESS.

Compressible flow. For compressible fluids such as gases and vapors it is necessary to add thermal terms to the Bernoulli equation. The energy equation between station 1 and station 2 then becomes

$$\frac{p_1}{w_1} + JI_1 + z_1 + \frac{V_1^2}{2g} - Jq = \frac{p_2}{w_2} + JI_2 + z_2 + \frac{V_2^2}{2g}$$

where q is Btu/lb of fluid flowing, which may be transferred from the fluid to the surroundings. If the heat flow is into the fluid, then q is negative, as the equation is written. Internal energy is thermal energy and is represented by I in Btu/lb, and J is 778 ft-lb/Btu. For a perfect gas $\Delta I = c_v \Delta T$ for each unit of weight. For a gas or a vapor the quantity p/w is usually large relative to $z_1 - z_2$, because of the small value of w in general. Therefore the z terms are usually, but not always, negligible.

Because p/w, which equals pv, and I are usually associated for gases and vapors, it is customary to replace them by a single term called enthalpy H, which may be described algebraically as $H = I + pv/J$ in Btu/lb (*see* ENTHALPY). Therefore the energy equation becomes

$$JH_1 + \frac{V_1^2}{2g} - Jq = JH_2 + \frac{V_2^2}{2g}$$

Isothermal flow of a gas. For a perfect gas $I_1 = I_2$ if the temperature is unchanged, and $p_1/w_1 = p_2/w_2$. Hence for this special case

$$V_2^2 - V_1^2 = -2gJq$$

which shows that an isothermal flow is accompanied by an absorption of heat when there is an increase in kinetic energy. This equation is true either with or without friction, because friction supplies some of the heat necessary to maintain constant temperature; however, less heat is absorbed from the surroundings, and q is less and the resultant increase in kinetic energy is less.

Adiabatic flow. In adiabatic flow, heat transfer is zero; hence the energy equation becomes

$$JH_1 + \frac{V_1^2}{2g} = JH_2 + \frac{V_2^2}{2g}$$

and because $778 \times 2g$ is practically 50,000, this may be written

$$V_2^2 - V_1^2 = 50{,}000(H_1 - H_2)$$

For a perfect gas $\Delta H = c_p \Delta T$ where c_p is specific heat per lb at constant pressure. Therefore for gases

$$V_2^2 - V_1^2 = 50{,}000\, c_p\, (T_1 - T_2)$$

Also for gases $c_p = kR/(k-1)\,J$, where $k = c_p/c_v$, the ratio of specific heats at constant pressure and at constant volume, and R is the gas constant. Then, because $pv = RT$,

$$\frac{V_2^2 - V_1^2}{2g} = \frac{k}{k-1}(p_1 v_1 - p_2 v_2)$$

These preceding equations are valid with or without fluid friction. For frictionless flow the values of H_2, T_2, and v_2 are all determined as the result of an isentropic expansion from p_1 to p_2. Fluid friction increases the numerical value of these quantities and thus decreases the numerical values of the right-hand sides of the equations. *See* ISENTROPIC FLOW.

Rotational and irrotational flow. In rotational flow, each minute particle of fluid rotates about its own axis. A specific case is a forced vortex in which a fluid rotates like a solid body. Such a case might be obtained with a fluid in a cylindrical vessel rotating about its central axis, assuming no motion of the fluid relative to the container. A more common case is that in which the container is fitted with vanes, as the impeller of a centrifugal pump. As a small element rotates about the central axis, it also rotates about its own axis (Fig. 7).

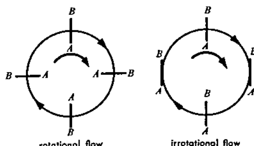


Fig. 7. In rotational flow, particles rotate on their axes as they move around; in irrotational flow, particles retain absolute orientation during motion.

In irrotational flow, each infinitesimal particle or element of the fluid preserves its original orientation. Because an element of fluid can be caused to rotate about its axis only by the application of viscous forces, rotational flow is possible only with a real or viscous fluid, whereas irrotational flow is possible only for an ideal or nonviscous fluid. For fluids of low viscosity, such as air or water, irrotational flow may be approached in a free vortex. In a free vortex, a body of fluid rotates without the application of any external torque, because of some angular momentum previously imparted to it. Examples are the rotation of fluid after leaving the impeller of a centrifugal pump, a tornado, or the rotation of water entering the drain of a bathtub. *See* VORTEX.

Boundary-layer flow. For an ideal frictionless fluid the velocity of flow adjacent to a surface would be the same as that at a distance. In actuality, adhesion between fluid and boundary surface tends to make the velocity of the fluid at the surface equal to the velocity of the surface body. For a very small distance from the surface, the velocity increases with distance at a very rapid rate because of viscosity within the fluid. The flow in this thin layer is laminar in character. This thin layer is known as the laminar boundary layer.

There is then a transition zone, the boundaries of which are indefinite, and beyond this the flow is fully turbulent. Much farther from the surface, the effect of the surface vanishes and the flow is undisturbed. The layer between the laminar one and the undisturbed field is known as the turbulent boundary layer. It is thick and there is no sharp line of demarcation between it and the undisturbed uniform flow. Such a case may be found with an airplane at a considerable height above the ground or a submarine deeply submerged beneath the surface. In the case of flow in a pipe or other conduit the turbulent boundary layers from opposite sides may meet so that there is no zone of undisturbed flow. Viscosity effects are most pronounced at or near a solid boundary and diminish rapidly with distance from the boundary. *See* FLUID-FLOW PRINCIPLES; FLUID MECHANICS; GAS DYNAMICS; HYDRODYNAMICS; *see also* FLOW MEASUREMENT. [R.L.DA.]

Bibliography: R. L. Daugherty and A. C. Ingersoll, *Fluid Mechanics*, 5th ed., 1954.

Flower (botany)

The flower is the characteristic reproductive structure of angiosperms, and it is interpreted morphologically as a specialized shoot. The flower consists of an axis bearing appendages, the flower parts or floral organs. Some of the floral parts—stamens and carpels—form reproductive structures; others—sepals and petals—are sterile. The sepals together constitute the calyx; the petals, the corolla.

Flowers develop from buds, as do other kinds of branches. *See* BUD (BOTANY). Plants may produce flowers and leaves in separate buds (elm, morning-glory); or both flowers and leaves may arise from the same, mixed buds (buckeye, apple).

FLOWER CHARACTERS

Flowers differ from vegetative shoots in that they show little or no elongation of internodes. *See* STEM (BOTANY). As a result, the different floral organs develop at close levels and are not distributed at conspicuous intervals along the axis, as are the leaves on branches of most plants. *See* LEAF (BOTANY). Also in flowers, buds usually do not develop in the axils ("armpits") of floral organs as they do in the axils of leaves of vegetative twigs.

The parts of a flower, like ordinary leaves, develop as lateral outgrowths of a growing point (apical meristem) of a bud. Often the lowermost floral organs develop first, followed in order by the others, with those nearest the tip of the bud generally emerging last. Other sequences occur also.

Flower sexuality. Flowers may be complete, having all four whorls of parts: sepals, petals, stamens, and carpels; or incomplete, lacking one or more whorls (Fig. 1). A flower that has both stamens and carpels is perfect (bisexual, hermaphroditic), regardless of whether sepals and petals are present. One carpel may constitute a pistil, or several carpels may be united into one pistil. If either stamens or carpels are absent, the flower is imperfect (unisexual); when stamens only are present, the flower is staminate (male); if carpels only are present, it is pistillate (female). Plants are monoecious if both staminate and pistillate flowers are produced on the same individual; dioecious, if staminate and pistillate flowers occur on separate individuals; and polygamous, if some perfect and some imperfect flowers occur on the same plant.

Wind-pollinated flowers (anemophilous) are usually inconspicuous, often without perianths (sepals and petals), whereas insect-pollinated flowers (entomophilous) have perianths, especially corollas, which are showy and often highly specialized.

Flower receptacle. The receptacle is the expanded end of the flower stalk (floral axis) from which the other floral organs arise (Fig. 2). In form, it is commonly flat or convex (strobiloid) but it may be concave or cup-shaped (cotyloid). Sometimes an outgrowth or expansion of the receptacle forms a cup-shaped structure called a hypanthium which bears sepals, petals, and stamens on its rim. The floral tube may resemble a

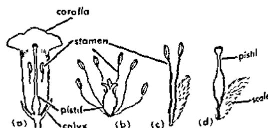


Fig. 1. Flower sexuality. (a) Perfect, complete, entomophilous. (b) Perfect, incomplete. (c) Staminate. (d) Pistillate, anemophilous.

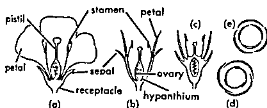


Fig 2. Flower receptacle and aestivation; arrangement of parts on receptacle. (a) Hypogynous (b) Perigynous. (c) Epigynous. (d) Imbricate. (e) Valvate.

hypanthium but it is formed by the coalesced and adnate (congenitally grown together) bases of the sepals, petals, and stamens. Where a common receptacle supports numerous flowers, as in a head, the individual flowers may be subtended by small appendages of the receptacle known as bracts (*see* INFLORESCENCE).

Insertion or adnation refers to the union of dissimilar parts of a flower, as of stamens and petals; if not so joined, they are said to be free. When all flower parts are attached to the receptacle below the pistil and free from it, the flower is hypogynous; if petals and stamens are inserted on the hypanthium or on the floral tube but free from the pistil, perigynous; if the floral tube is adnate to the ovary (lower, ovule bearing part of pistil), epigynous.

Aestivation, or praefloration, refers to arrangement of parts, particularly of calyx (sepals) and corolla (petals), in the bud. If sepals or petals overlap like shingles, they are imbricate; if they lie edge to edge, they are valvate.

Flower parts. The number of parts in each set may be indicated by the ending -merous with a prefix, as pentamerous, having parts in fives or multiples of five (Fig. 3). During development, if members of one set become fused with another set, they are said to be adnate or adherent; if the parts of a whorl become fused, the terms symphysis, coalescence, connation, and cohesion are applicable. The term distinct indicates complete separation of like parts; free denotes separate unlike parts. In the ray flowers of composites, the corolla forms a strap-shaped structure called a ligule.

Calyx. The outermost whorl, composed of sepals, which may be distinct or connate, is called the calyx. If connate, the calyx is gamosepalous or synsepalous. Prominent veins of the calyx, especially of the tubular type of calyx, are called nerves. In some families, the calyx has the form of bristles; such calyx is called pappus.

Corolla. The second whorl of flower parts, composed of petals, is called the corolla, which in many species contributes to the showiness of the flower. If the corolla is wanting, flowers are described as apetalous; if petals are distinct, the corolla is polypetalous; if petals are fused, the corolla is gamopetalous or sympetalous. The calyx and corolla constitute the perianth, or floral envelope. Flowers with petals alike in size and shape are regular, radially symmetrical, and actinomorphic. If the petals are unlike in size or shape, the flowers are irregular, usually bilaterally symmetrical and zygomorphic. The term symmetrical, in a taxonomic sense, implies an equal number of parts in each series of the perianth.

Stamens. A stamen consists of an anther (pollen-producing structure) and a filament or supporting stalk (Fig. 4). The collection of stamens is called the androecium. Stamens may be arranged in one or more whorls or along a spiral. Anthers are oriented in different ways: versatile, if attached near the middle enabling them to swing freely; introrse, if they face inward toward center of flowers; extrorse, if they face outward. Stamens are usually distinct, but filaments may be united into one set (monadelphous) or into two sets (diadelphous). A stamen with no functional anther is a staminode and may appear glandlike.

Pistils A pistil is either simple, composed of one carpel; or compound, consisting of two or more

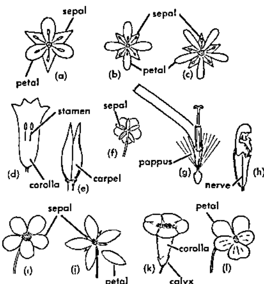


Fig 3. Flower parts. (a) Trimerous (b) Tetramerous. (c) Pentamerous. (d) Adnation, stamens on corolla. (e) Coalescence, carpel to carpel. (f) Sepals distinct. (g) Ligulate corolla and pappus. (h) Synsepalous. (i) Apetalous (sepals petal-like). (j) Polypetalous and regular. (k) Sympetalous (l) Irregular.

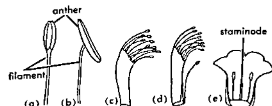


Fig. 4. Stamens (a) A stamen. (b) Versatile anther. (c) Monadelphous. (d) Diadelphous. (e) Flower opened to show staminode.

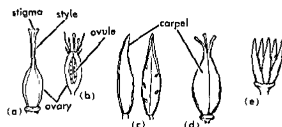


Fig 5 Pistils (a) Gynoecium syncarpous, ovary superior. (b) Ovary inferior. (c) Simple, pistil carpel closed and open. (d) Compound pistil. (e) Gynoecium apocarpous.

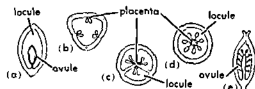


Fig. 6. Placentation. (a) Parietal (1 carpel). (b) Parietal (3 carpels). (c) Axile (3 carpels). (d) Free central (5 carpels). (e) Free central, longisection.

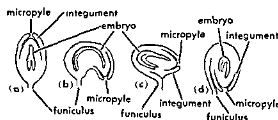


Fig. 7. Ovule types. (a) Orthotropous. (b) Campylotrophous. (c) Amphitropous (d) Anatropous.

carpels (Fig. 5). Gynoecium is the technical term for the collection of carpels. If carpels are distinct, the gynoecium is apocarpous; if two or more carpels are joined together, the gynoecium is syncarpous. A pistil has three parts: stigma, style, and ovary. The ovary, usually an enlarged base, contains the ovules; the stigma is the receptive surface for pollen, and the style connects the stigma and ovary. An ovary is considered superior if it is free from the other parts (hypogynous and perigynous flowers); inferior, if the receptacle or floral tube is adnate to the wall of the ovary (epigynous flower).

Placentation. The placenta is the region of the ovary to which ovules are attached, usually the margin of the carpel (Fig. 6). Kinds of placentation vary with the structure of the ovary: parietal, if the placentas occur on the wall of the ovary; axile, if each carpel forms a separate locule (cell) with the placentas at center, free central, if ovules are borne on a central axis which extends upward from the base of a one-celled ovary.

Ovules. An ovule is a rudimentary seed (Fig. 7). In the development of ovules ...
sul
fur

the body of the ovule bent or curved; amphitropous, ovule symmetrical but half inverted with the micropyle and funiculus at right angles to each other; anatropous, inverted, one side adnate to the funiculus with the micropyle close to the attachment of the funiculus. *See* PLANT TAXONOMY.

[N.A.]

FLOWER ANATOMY

Morphologically the flower is regarded as a contracted shoot (leafy stem of determinate growth) with shortened internodes and crowded appendages (Fig. 8). It arises from a floral growing point (apical meristem) that shows a more or less different structure than the apical meristem of a leafy stem (*see* MERISTEM, APICAL). A floral growing point results either by gradual or abrupt transformation of a vegetative apex.

Development. Primordia of floral organs develop laterally either in a compact spiral or a series of cycles (whorls) on the floral apical meristem, usually at successively higher levels, that is, in an acropetal or centripetal sequence (Fig. 9). Most floral organs are fully developed when the flower opens; however, the carpels remain immature until the fruit ripens.

Vascular system. All floral organs are supplied with vascular traces (vascular bundles pertaining to lateral organs) from the vascular system of the stem, which may be simple or complex at this point (*see* VASCULAR BUNDLES). Usually the number of traces per floral organ is constant, but generally it is an odd number (one, three, and so on).

Peduncle. The stalk of an entire inflorescence is commonly known as the peduncle. The individual flower stalks (pedicels) are inserted on it.

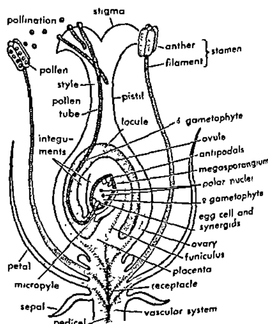


Fig. 8. Structures of a flower in median longitudinal section.

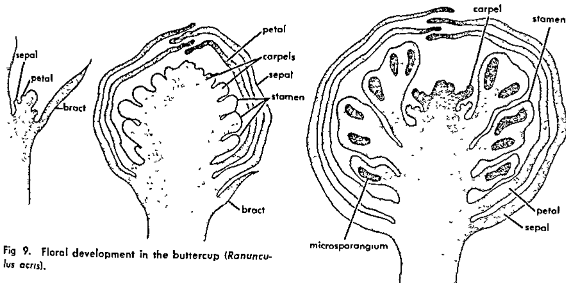


Fig 9. Floral development in the buttercup (*Ranunculus acris*).

Pedicel. The stalk of a single flower is called a pedicel. It represents the last internode of a branch which ends in a flower. The vascular tissue in the pedicel appears either as an unbroken cylinder or a system of interconnected strands.

Receptacle of the flower. Commonly the apex of the pedicel is enlarged forming the floral receptacle or thalamus (also called torus). It bears all floral organs and its vascular stele supplies them with vascular traces. The latter diverge from the receptacular vascular system either in whorls or in spirals. Some bundles, however, may continue beyond the last floral organs and terminate near the top of the receptacle. The size and shape of the receptacle determine in large measure the arrangement and position of the floral organs as well as the appearance of the flower. The receptacle may be conical or flattened and variously expanded (into a disk) or depressed and even flask-shaped enclosing an inferior ovary. Or the receptacle may be greatly elongated between some cycles of floral organs.

Sepal. The outermost set of floral organs, if present, consists of sepals, which are usually green (contain chlorophyll) and are small. Collectively they form the calyx. They may be free or united (synsepalous). Their vascular supply is much like that of ordinary leaves.

Petal. Above or interior to the calyx are the petals, collectively called the corolla. They are often brightly colored, variously shaped, and free or fused (sympetalous). They may or may not be present. Normally each petal receives a single vascular trace.

Perianth. The calyx and corolla together are referred to as the perianth or perigone. Dichlamydous flowers have both calyx and corolla. Flowers are heterochlamydous if calyx and corolla are different in structure. Homoiochlamydous flowers have perianth leaves that are alike, that is, not differentiated into sepals and petals. The members of such perianths are called tepals. Monochlamydous flowers possess only one set of floral envelopes, whereas achlamydous flowers lack both sets.

Stamen. Compared with the accessory organs (sepals and petals), stamens and carpels are regarded as the essential organs of the flower. The stamen is most commonly interpreted as a microsporophyll (leaflike structure) bearing microsporangia. Normally it consists of a stalklike sterile part, the filament, and a fertile part, the anther. The central part of the filament bearing the anther is known as the connective tissue since it unites the two parts of the anther, called thecae, each of which contains two pollen sacs (microsporangia). Anthers are basifixed when the connective tissue forms a direct continuation of the filament, dorsifixed when the filament is attached to the abaxial (away from the axis) side of the connective, or apicifixed when the filaments are attached to the upper end of the connective. If the filament is attached by a single point to the connective, anthers are versatile and they can swing freely as in wind-pollinated grasses.

The vascular supply of most stamens is provided by a single-trace connection with the floral vascular system. In relatively primitive flowering plants, a three-trace vascular supply is found as also occurs in their carpels and leaves. Some primitive groups have undifferentiated broad, strap-shaped stamens, whose microsporangia are adnate on the inner surface, as in the plant families Winteraceae and Himantandraceae. Otherwise stamens may assume many shapes and sizes and may fuse with one another (synstemony) or with the petals or corolla tube (epipetaly), or with sepals (episepaly). Monadelphous stamens occur in a single group, whereas diadelphous stamens form two groups. Heterostemony refers to the presence in the same flower of two or more different forms of stamens as in the Leguminosae and Labiatae. Staminodes are vestigial stamens. Introrse anthers open by longitudinal splits, pores, or other openings on the inner

(adaxial) side, whereas extrorse anthers open on the outside (abaxial) side.

The development of the anther begins with a homogeneous mass of meristematic tissue that becomes bilobed and later quadrilobed, each lobe representing a microsporangium. Then certain cells of the subepidermal layer in each lobe become larger and their walls thicker while their protoplasm becomes denser than that of neighboring cells (*see* EPIDERMIS PLANT; PROTOPLASM). These modified cells (one to many) function as the initials of the archesporium. Periclinal (parallel to surface) divisions of these initials separate to the outside the so-called primary parietal cells from the inner primary sporogenous cells. By subsequent periclinal divisions, the parietal cells form several concentric wall layers. Meanwhile the sporogenous cells divide and form a solid mass of cells. Later the individual sporogenous cells separate and round off, becoming pollen (microspore) mother cells (Fig 10). Each pollen mother cell then undergoes meiosis and gives rise to a tetrad of pollen grains (*see* MEIOSIS). Meanwhile in each anther lobe a wall-zone, three to four cells thick, is being formed by tangential divisions of the parietal and other cells adjacent to the sporogenous tissue. Thus, the four pollen sacs, with their sporogenous tissue enclosed in wall layers, become delimited from the central tissue of the connective. The epidermis of the pollen sac wall is derived from the outermost layer of cells of the young anther. The subepidermal layer is the endothecium. The anticlinal (perpendicular to the surface) walls of endothecial cells become thickened with bands of lignified cellulose arranged in patterns characteristic throughout whole families. Later the protoplasm disappears in these cells. The endothecium may be confined to walls of the anther lobes or may surround the whole anther. In addition to the endothecium, several more wall layers are formed—one or more middle layers and the innermost layer, or tapetum. Commonly the middle layers are completely crushed. The cells of the tapetum divide and elongate centripetally while the pollen mother cells are being formed. Often the nuclei increase in number and change their normal shape. Eventually the walls dissolve more or less

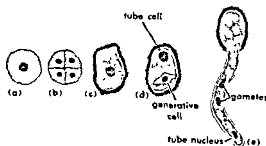


Fig. 10. Development of pollen. (*a*) *Microspore mother cell.* (*b*) *Tetrad of microspores (pollen grains).* (*c*) *Pollen grain* (*d*) *Pollen grain with generative cell and tube cell.* (*e*) *Germination of pollen; pollen tube contains tube nucleus and two gametes.*

completely. In some species the protoplasts also become disorganized; in others the liberated protoplasts form a united plasmodium that intrudes among the pollen mother cells. As the pollen ripens, the plasmodium disappears. The sporangial wall between two adjacent lobes is removed so that the two pollen sacs on each side of the anther are united. The anther is now bilocular. Most of the stages described to this point may occur while the flower is still in a closed bud. When the anther is exposed to air and light, the sporangial walls lose water and contract. In the endothecium the unthickened inner and outer walls contract more readily than the thickened anticlinal walls. This change results in dehiscence (splitting), outward curling of sporangial walls, and discharge of pollen. Anthers usually open gradually, and not explosively.

Pollen. The appearance of pollen mother cells marks the end of the sporophyte stage and is often followed by a resting period lasting from a short pause in annuals to several months in woody species in which the flowers open in the second season after their initiation. As pollen mother cells undergo meiosis, tetrads are formed, usually while still enveloped by the mother cell wall. Each cell of the tetrad then develops its own special wall. Although some families (Orchidaceae, Naiadaceae) produce pollen with only slightly thickened walls, most pollen grains develop two distinct walls, an exine and an intine. The exine soon thickens considerably and develops characteristic spines or other appendages, and has pores or furrows where the exine is thin and where later the pollen tubes emerge. Like the walls of fungal spores, the exine of pollen grains contains a substance called sporopollenin, a substance related to suberine and cutin but more resistant than these (*see* FUNGI). The sporopollenin seems to help preserve pollen found in peat and older deposits, and is thus of considerable importance in sciences utilizing pollen for plant identification, such as paleobotany, plant taxonomy, and plant ecology (*see* PEAT). The intine varies considerably in thickness, being thickest under the pores. Its innermost layer contains cellulose, the outer one pectin. The intine has great affinity for water and swells readily, especially below the pores, where it eventually ruptures the exine and protrudes.

Pollen grains are rich in carbohydrates and fats, and may contain pigments (anthocyanins, but not chlorophyll) and aromatic substances. Pollen grains are variously shaped and sculptured. Monocotyledonous pollen grains are in bilateral arrangement within the mother cell and are commonly boat-shaped and single-furrowed (monocolpate) at maturity. Dicotyledonous pollen grains are formed in a tetrahedron (providing three surfaces of contact with the other grains) and are typically three-pored or three-furrowed (tricolpate). Pollen of wind-pollinated plants is usually smooth and dry. Insect-pollinated plants often have oily coatings on their pollen. Pollen grains may vary in size from several microns (μ) to gigantic forms larger than

200 μ. The number of grains per pollen sac (locule) may vary from approximately 30 (largest sized grains) to more than 50,000 (about 5μ per grain). Sometimes grains remain attached to others forming small groups called massulae, or the grains of an entire locule remain attached and form a pollinium as in Orchidaceae.

Pollen grains are mostly uninucleate until they separate from the tetrads. This stage is usually followed by mitosis, during which a generative nucleus is cut off toward the pole which was innermost in the tetrad stage (*see* MITOSIS; REPRODUCTION, PLANT). This generative nucleus later gives rise to the two male gametes (sperms), whereas the other nucleus becomes the vegetative nucleus of the pollen grain and later, after the pollen germinates, constitutes the tube nucleus. With the first nuclear division the pollen grain becomes the male gametophyte.

After mitosis, the generative nucleus stains densely and contains considerable desoxyribonucleic acid (DNA). The vegetative nucleus loses most of its staining capacity, whereas its surrounding cytoplasm gains in staining capacity and is rich in protein and ribonucleic acid (RNA), factors favoring further growth. A delicate membrane appears between the two nuclei so that the generative nucleus then lies in a small, lenticular cell close to the wall of the grain. Eventually it becomes free as a biconvex cell, but remains smaller than the vegetative cell. Vacuoles disappear and starch grains or oil drops may accumulate. The pollen grain may now be regarded as mature. The generative nucleus may divide before the pollen germinates or later in the pollen tube. The vegetative nucleus as a rule does not divide. By an extension of the intine of the pollen grain through a furrow or a pore, the pollen tube is formed and proceeds to grow at its apex. The pollen tube is smooth-walled and cylindrical and the entire contents of the pollen grain pass into it. The vegetative nucleus enters the tube first and thereafter is called the tube nucleus. The male nuclei, produced by division of the generative nucleus, often change their shape and become vermiform (worm-shaped), especially when they enter the embryo sac. The wall of the tube is composed of cellulose but its tip apparently consists of more plastic materials. After having grown a long distance toward their ultimate destination, the embryo sac, the pollen tubes may assume a volume hundreds of times larger than that of the original grain. Nutrients for pollen-tube growth are obtained from starch grains received from the original pollen grain and probably also from cells of the nutritive tissue in the style, with which the growing tube comes in contact. The rate [illegible]

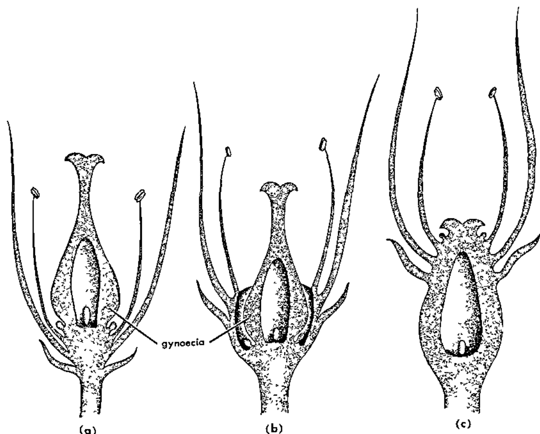


Fig 11. Position of gynoecia. (a) Superior; (b) half-inferior; (c) inferior.

tal conditions, especially temperature, markedly affect the rate of growth. Pollination (transfer of pollen from an anther to a stigma) may precede fertilization (union of sperm and egg) by minutes, hours, days, and even months Normally each pollen grain emits only one pollen tube. If several tubes are formed only one becomes functional.

Carpel. The term carpel is applied to the megasporophyll (megasporangia-bearing leaf) of angiosperms. Carpels represent the inner essential organs of the flower. Since carpels form more or less completely closed structures, the ovules (and later seeds) are concealed within them Flowers may have one to many carpels. The single carpel or the carpels collectively, if more than one are present, are referred to as the gynoecium. The gynoecium occupies the center or apex of the flower and thus terminates the growth of the receptacle. In regard to position of the gynoecium, three types of flowers are usually recognized (Fig. 11): a hypogynous flower, characterized by the insertion on the receptacle of stamens, petals, and sepals below the carpels (the gynoecium is superior); a perigynous flower bearing the floral organs on the rim of an expanded saucer- or cup-shaped receptacle or floral tube (the gynoecium is half-inferior); an epigynous flower, having the carpels deeply imbedded in a flask-shaped receptacle, or floral tube, with the other floral organs inserted above, at the rim of the flask (the gynoecium is inferior). Because the receptacle or floral tubes are adnate to an inferior gynoecium, it is difficult to distinguish the constituent parts of such a gynoecium except by a developmental anatomical study.

A single carpel may form a simple pistil or several may be united into a compound pistil. Thus the gynoecium is either apocarpous, composed of one to many separate carpels or simple pistils, or syncarpous, composed of two or more concrescent (fused) carpels (compound pistil).

Pistils generally consist of an ovary, the lower flask-shaped part containing the ovules (one or more): a style, the narrow necklike sterile extension of the ovary; and a usually terminal and often spherical stigma, the receptive structure on whose adhesive hairy surface pollen grains land and germinate.

In a simple pistil the carpel may be pictured as a folded leaf with the margins turned toward the center of the flower (adaxially) and united. Concrescence of carpels into a compound pistil may be partial or complete, varying from ovaries fused only at their base, through those with united ovaries but free styles, or only free stigmas, to those with fused styles and stigmas. A compound ovary may be unilocular if the carpels are united margin to margin. It may be plurilocular if the carpels are united in a folded condition. In such an ovary the united sides of the concrescent carpels form partitions, or septa (true dissepiments), that divide the space within the ovary into a number of separate cavities or locules. A basically plurilocular ovary may become unilocular by the evolutionary disappearance of the septa. Septa consisting of outgrowths of midribs or other parts of carpels are called false dissepiments. The style contains a tissue that seems to facilitate growth of the pollen tube and provide it with food. Thus this tissue resembles the stigma in its function and therefore is called stigmatoid tissue. Some styles are hollow and the canal-line cavity is lined with stigmatoid tissue.

Carpels may be connected with the vascular tissue of the receptacle by one, three, five, or more vascular traces, most commonly three. The median trace is continued as the dorsal bundle in the carpel and forms the midrib. The two outermost bundles are called ventral or marginal bundles because they extend along or near the edge of the carpel All these bundles may be branched. The three-trace connection with the receptacle is generally regarded as primitive. Bundles usually extend to the base of the ovule and continue into the latter through its stalk, or funiculus.

Placenta is the generally used term for the region within the ovary that bears the ovules. Placentation refers to the arrangement of ovules within the ovary (Fig. 12). In regard to the ovary, placentation may be parietal, ovules located on the ovary walls, or central, ovules located in the center; the latter may be axile, free central, or basal. In regard to the carpels themselves, placentation may be marginal if ovules occur near the margins of carpels, or laminal if they are situated on the inner surface of carpels.

Ovules may be sessile or stalked, that is, borne on a funiculus (funicle), and may assume various positions within the ovary. They are either erect (orthotropous, or atropous), horizontal, ascending or pendulous, or suspended (inverted, or anatropous). Half-inverted ovules are known as amphitropous and those with a bent body are called campylotropous. As ovules mature, characteristic structures may appear such as an aril or strophiole formed by the funiculus as a small protuberance on the seed, or a raphe marking the original line of union between funiculus and ovule. *See* SEED (BOTANY).

Dicotyledonous ovules are either unitegmic or bitegmic; that is, they are enveloped by one or two integuments. Many dicotyledons have bitegmic ovules, a condition generally regarded as primitive Single integuments apparently result from the fusion of two or the loss of one, and are found in most sympetalous orders, also in Betulaceae, Juglandaceae, and Umbelliferae (*see* ARCHICHLAMYDEAE; SYMPETALAE). Parasitic families, such as Santalaceae, lack integuments. Integuments usually enclose ovules on all sides, leaving only a small opening at the tip, the micropyle, through which the pollen tube commonly enters. The base of the ovule, situated directly opposite the micropyle, is called the chalaza.

The interior of the ovule is occupied by an oval central mass of tissue called the nucellus (megasporangium). The nucellus contains the megagametophyte, or embryo sac (Fig. 13). When mature, the embryo sac frequently contains seven cells and eight nuclei characteristically arranged as follows:

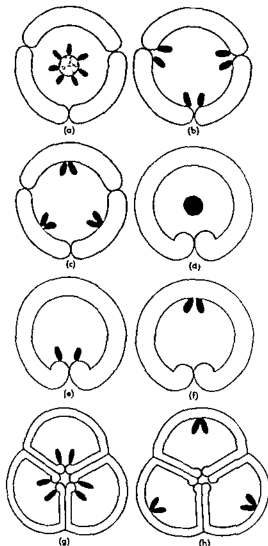


Fig 12. Types of placentation; ovules are shown in black. (a,d) Free central placentation. (b,e) Parietal, marginal placentation. (c,f,h) Parietal, laminal placentation. (g) Axile, marginal placentation.

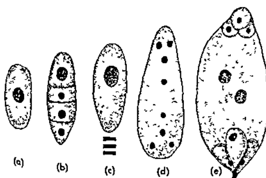


Fig 13. Development of embryo sac. (a) Megaspore mother cell. (b) Tetrad of megaspores. (c) Surviving megaspore. (d) Embryo sac with eight nuclei. (e) Mature embryo sac, showing egg cell and synergids, polar nuclei, and antipodals.

three cells, known as the egg apparatus, are found near the micropylar end, the central one being called the egg, the other two, synergids; the center of the embryo sac is occupied by a large cell with the two polar nuclei; the remaining three cells, or antipodals, are grouped together at the chalazal end. The polar nuclei may fuse into the *so-called fusion* nucleus or primary endosperm nucleus. If the nucellus is composed of a massive tissue, which may even remain as a nutrient tissue in the ripening seed, the ovule is defined as crassinucellate If the nucellus is small, sometimes only a single layer of cells in thickness, the ovule is called tenuinucellate.

Early in the growth of the ovule, the development *of the embryo sac (female gametophyte) is initi*ated by the differentiation of an archesporial cell, a subepidermal cell in the micropylar region that becomes conspicuous because of its size and dense contents. *The archesporial cell may form a parietal* (wall) cell and a megaspore mother cell or may function directly as a megaspore mother cell. Sometimes more than one megaspore mother cells are formed In most flowering plants, the megaspore mother cell undergoes meiosis consisting of two successive nuclear and cell divisions. Usually four haploid (with half the somatic number of chromosomes) megaspores, arranged in linear fashion, result from this process. As a rule, the three megaspores nearest the micropyle disintegrate and only the chalazal one becomes functional, giving rise to a single female gametophyte. Generally three successive free-nuclear divisions result in the formation of eight nuclei. These are separated from each other by walls in such a way that the seven-celled structure characteristic of a mature embryo sac is formed An embryo sac, derived from a single functional megaspore, is known as a monosporic *embryo sac*. In some plants the megaspore mother cell fails to divide into megaspores. Instead, it forms directly an embryo sac, a so-called tetrasporic embryo sac, by nuclear division. Altogether *10 major types of formation of embryo sacs have* been described and named after the genera in which they occur, such as, for example, *Polygonum*, *Oenothera*, *Allium*, and *Adoxa*.

Normally pollen tubes enter the embryo sac through the micropyle, a mode called porogamy. If entry is gained elsewhere, the process is called aporogamy.

Fertilization. *Although several pollen tubes may* enter the embryo sac, only the nuclei of one participate in fertilization (syngamy). It is a double fertilization, for one of the generative nuclei (sperm nuclei) fuses *with the egg, and the other with the* polar nuclei. The fusion of the sperm nucleus with the polar nuclei or the product of their fusion is called triple fusion. The fertilized egg forms the zygote *(first cell of the diploid sporophytic em*bryo) which is surrounded by a delicate membrane. The triploid nucleus resulting from the fusion of the two polar nuclei with one sperm nucleus becomes the endosperm *nucleus. The divisions of* a nucleus *and its products give rise to a tissue*

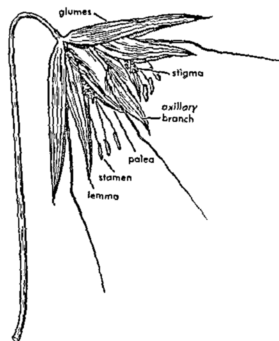


Fig. 14. Spikelet of wild oat (*Avena fatua*).

endosperm, usually filled with nutrient substances needed by the developing embryo.

Nectary. Many flowers contain specialized organs or modifications (glandular surfaces) of floral organs, called nectaries (glands), secreting sugar-containing nectar (*see* SECRETORY STRUCTURES, PLANT). The presence of nectaries is essential for successful pollination by insects. According to position in the flower, nectaries may occur on the receptacle, in and around ovaries, on stamens, or on the perianth. They may be in exposed positions on the flower or deeply imbedded in the ovary and more or less concealed.

Abscission. Floral organs such as petals, stamens, and other parts may drop off at various times, either before or after wilting. Usually a separation layer is formed shortly before a fl...

... much simpler (*see* PLANT GROWTH).

Floret. This is a term applied to a flower of a group. The florets are usually borne singly or in groups of two or more in compact units called spikelets, which in turn are variously combined into inflorescences (Fig. 14). A *spikelet* is a short shoot, whose axis, the rachilla, bears a number of reduced leaves (bracts, or glumes) and reduced axillary branches which develop into florets. The lowest bracts are called glumes, whereas the bracts borne above them are known as lemmas. The axil of each lemma bears a flower which is normally enclosed on the opposite side by another bract, palea. Lemma, palea, and flower together constitute a floret. Glumes and lemmas may be sheaths of modified leaves, and often bear long, bristlelike awns. Essentially constructed like other flowers, grass flowers contain, in addition to reproductive organs, two (sometimes three) lodicules, which are colorless scales, usually thickened at the base and supplied with vascular tissue. During flowering time they fill with water and become two or three times larger than before. Thereby the floret is opened and the floral organs are exposed. *See* GRASS CROPS.

[T.K.J.]

Bibliography: *See* PLANT ANATOMY.

Fluid coupling

A device for transmitting rotation between shafts by means of the acceleration and deceleration of a hydraulic fluid. Structurally, a fluid coupling consists of an impeller on the input or driving shaft and a runner on the output or driven shaft. The two contain the fluid (Fig. 1). Impeller and runner are bladed rotors, the impeller acting as a pump and the runner reacting as a turbine. Basically, the impeller accelerates the fluid from near its axis, at which the tangential component of absolute velocity is low, to near its periphery, at which the tangential component of absolute velocity is high (Fig. 2). This increase in velocity represents an increase in kinetic energy. The fluid mass emerges at high velocity from the impeller, impinges on the runner blades, gives up its energy, and leaves the runner at low velocity (Fig. 3).

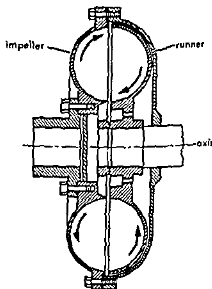


Fig. 1. Basic fluid coupling.

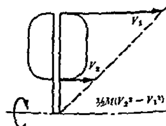


Fig. 2. Hydrokinematic principle.

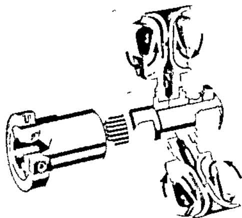


Fig 3. Sheave with fluid coupling. Dark portion is driven directly by motor; light portion drives the load. (*Twin Disc Clutch Co.*)

Coupling characteristics. A fluid coupling absorbs power from an engine by means of the centrifugal pump or impeller, which increases the moment of momentum of the mass of fluid. As a consequence, the input power required to drive the coupling increases as the cube of the speed and as the fifth power of the outside diameter of the impeller. Similarly, the capacity of the coupling to absorb torque varies as the square of its input speed.

Because the efficiency of a coupling is a function of slip, couplings normally operate at 95–98% efficiency. The fluid follows a continuous toroidal helix, the pitch of which depends on the slip or relative difference in rotational velocity between impeller and runner. Some slip is necessary for energy transfer. As the slip increases at a given input speed, the output shaft turning slower than the input shaft, the quantity of fluid pumped between impeller and runner increases, with consequent increase in torque absorption and transmission. However, because the fluid flow cannot change abruptly, a fluid coupling absorbs rather than transmits shock loads from output to input. Torsional vibrations are similarly absorbed

A fluid coupling is designed for a normal running load at 2–5% slip. Conservation of momentum requires that, although shaft speeds may differ, input and output torques must be equal. Thus, for example, in starting a machine from standstill, with the output from the fluid coupling locked, the engine or motor is completely unloaded initially. As the engine starts up toward running speed, the drag torque characteristic of the fluid coupling lies below the developed torque characteristic of the engine. The engine can reach running speed faster with consequent saving in fuel or electrical consumption. The slip of the fluid coupling enables the engine to operate near its maximum-torque speed while accelerating the load, although this characteristic differs from that of a variable-torque transmission. *See* TORQUE CONVERTER.

Typical application. The effect of a fluid coupling can be described in terms of a specific application. If an electric motor is starting with a high inertia load that requires nearly full-motor torque, the motor accelerates slowly when connected directly to its load. However, with fluid coupling, the motor accelerates rapidly along A to C in Fig. 4. Meanwhile, the torque-transmitting capacity of the coupling builds up along B to C until it equals the developed motor torque. The motor is then able to deliver well over full load torque to accelerate the load, the torque demand falls from C to D as the load comes up to speed D. Meanwhile, the motor further accelerates C to E until steady state is reached with coupling slip represented by differential speed D to E. Use of the fluid coupling improves performance by faster load acceleration and

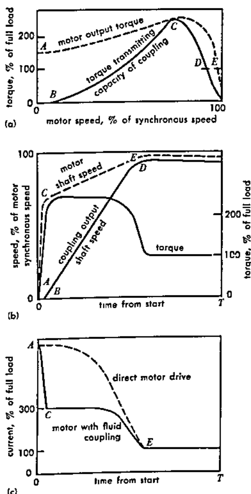


Fig. 4 Starting characteristic of electric motor with fluid coupling. (a) Torque speed during acceleration. (b) Acceleration characteristic (c) Current characteristic.

results in appreciably less power loss due to motor heating. *See* TRANSMISSION, AUTOMOTIVE. [H.J.WI.]

Bibliography: R. L. Daugherty and A. C. Ingersoll, *Fluid Mechanics*, 5th ed., 1954.

Fluid dynamics

The science of fluids in motion. Fluid dynamics attempts to describe the motion of a fluid as it is displaced and deformed by the action of moving or fixed boundaries. Fluid dynamics may be divided into two parts, hydrodynamics and aerodynamics. For low Mach numbers (ratio of velocity of fluid to local acoustic velocity) both hydro- and aerodynamics may be treated in the same manner, but for Mach numbers over about one-half, compressibility becomes important and must be taken into account.

Fluid dynamics makes use of both theoretical developments and experimental results. Simplifying assumptions are generally made in theoretical studies in order to make the equations manageable. The extent to which the results differ from the flow of real fluids must be determined by experiment, and corrections applied to the theoretical treatment, to obtain practical results. *See* FLUID MECHANICS.

Theoretical hydrodynamics has been made a useful tool in solution of flow problems by use of the Prandtl hypothesis, which states that with fluids of low viscosity the effects of viscosity are limited to a narrow region along the boundaries. The problem may be solved as if the fluid were frictionless (nonviscous) to determine the velocity and pressure intensity throughout the fluid except at the boundaries (*see* HYDRODYNAMICS). The flow near the boundaries is called boundary-layer flow, and takes into account viscous effects and the fact that the velocity at the boundary relative to the boundary is *zero*. From a study of the boundary layer, its growth may be computed, as well as the tangential shear force it exerts on the boundary. Under certain conditions of adverse pressure distribution, the boundary-layer film immediately adjacent to the wall comes to rest, and the bounding streamline *separates from the wall*. This phenomenon is known as separation, and results in turbulence and formation of a wake downstream from the separation point, which increases the energy losses. *See* BOUNDARY-LAYER FLOW. Beyond the separation point, the fluid does not follow the boundary and does not regain pressure as *velocity* is reduced. There remains an additional drag or pressure force on a body immersed in a flowing fluid, known as form drag.

At very high altitude or low pressure, instead of the Prandtl hypothesis, it is assumed that individual molecules slip along the *surface of a flight* vehicle in slip flow (*see* SUPERAERODYNAMICS). If the fluid is ionized and reacts with electric or magnetic fields as well as with *thermal and* mechanical boundary conditions, additional behavior is obtained [illegible]

[illegible]

19[illegible]

Fl[illegible] *Fluid Dynamics*, 1958.

Fluid mechanics

The science concerned with fluids, either at rest or in motion. It deals with pressures, velocities, and accelerations in the fluid, including fluid deformation and compression or expansion. Fluid mechanics may be divided into two branches, fluid statics and fluid dynamics, the first dealing with pressure intensities and forces exerted by a fluid at rest, and the second with forces exerted on fluids and their resulting motions. *See* FLUID DYNAMICS; FLUID STATICS.

The laws of fluid mechanics control a great portion of natural phenomena. The flight of an insect or bird, the motion of a fish through water, the relative movement of air masses as in frontal weather systems, and the eruption of a volcano are examples of flow that follow the laws of fluid mechanics. It is involved in many phases of aeronautical, chemical, civil, and mechanical engineering. The science of fluid mechanics requires the combination of theoretical analysis and orderly experimentation. In the theoretical approach, assumptions are made as needed to keep the resulting mathematical expressions manageable, such as assuming an incompressible fluid, or one without viscosity. The experimental work must be planned in such a way that similitude relations are used to a maximum advantage. *See* MODEL THEORY.

Definitions. Flow may be classified in many ways, such as turbulent, laminar; real, ideal; isothermal, isentropic; steady, unsteady; and uniform or nonuniform.

Turbulent-flow situations are most prevalent in engineering practice. The fluid particles move in irregular paths, causing an exchange of momentum from one portion of fluid to another. The fluid regions involved in the transfer of momentum due to turbulence can range in size from large-scale turbulence with many cubic miles of fluid participating in a single tornadic eddy to a small-scale turbulence with only a few thousand molecules interacting (*see* TURBULENT FLOW). Turbulent flow causes the conversion of mechanical energy into thermal energy at a rate varying roughly as the square of the *velocity*.

In laminar flow, particles move along smooth paths in layers, or laminae, with one layer gliding over an adjacent one. Laminar flow is governed by Newton's law of viscosity (*see* NEWTONIAN FLUID). It may be considered as flow in which all turbulence has been damped out by the action of viscosity.

A real fluid always has viscosity and, whether a liquid or a gas, has compressibility. An ideal fluid is considered to be frictionless (nonviscous) and incompressible. There is no means by which an ideal fluid can convert mechanical energy into thermal energy. When ideal fluid particles *have no* rotation, the fluid is irrotational and a velocity potential exists. *See* FLUID-FLOW PROPERTIES.

When gas flows without change in temperature, the flow is isothermal. When flow occurs such that no heat is added or subtracted at the boundaries,

the flow is adiabatic. Reversible adiabatic (frictionless adiabatic) flow is called isentropic flow.

Steady flow occurs when conditions at any point in the fluid do not change with the time. In unsteady flow, one or more quantities such as density or velocity change with time at a point. Uniform flow occurs when the velocity vector throughout the fluid is everywhere the same at any instant, and nonuniform flow occurs when the velocity vector has varying values throughout the fluid at any instant. The strict definitions of steady flow and uniform flow are relaxed in practical flow situations, due to turbulent fluctuations in the first case and to variations over a cross section in the second case. For example, liquid flow through a long pipe at constant rate is steady-uniform flow; liquid flow through a long pipe at a decreasing rate is unsteady-uniform flow; flow through an expanding tube at a constant rate is steady-nonuniform flow; and flow through an expanding tube at an increasing rate is unsteady-nonuniform flow.

A streamline is a continuous line through the fluid that has the direction of the local velocity vector at every point. A stream tube is composed of all streamlines through a small closed curve

Basic equations of fluid flow. In any fluid flow situation, three conditions exist (1) Newton's laws of motion hold for every particle at every instant (*see* EULER'S MOMENTUM THEOREM), (2) the continuity relationship holds; that is, net mass inflow into any small volume per unit time equals its time rate of increase of mass (*see* HYDRODYNAMICS), (3) at a boundary, the velocity component normal to the boundary equals the velocity component of the boundary normal to itself. For real fluids, in addition, the tangential component of fluid velocity at the boundary is zero relative to the boundary.

Integration of Newton's second law of motion may lead either to the Bernoulli equation or to the momentum equation. By considering the steady flow of a frictionless, incompressible fluid, the Bernoulli equation states that the mechanical energy remains constant along a streamline. By considering steady, irrotational flow of a frictionless, incompressible fluid, a form of the Bernoulli equation is obtained that shows that the energy is constant everywhere throughout the fluid (*see* BERNOULLI'S THEOREM). For steady flow, the momentum equation states that the resultant force acting on any free body of fluid is just equal to its time rate of change of momentum or, for a fixed control volume, that the resultant force equals the difference between the momentum per unit time leaving and the momentum per unit time entering.

The continuity equation for steady flow states that the mass per unit time flowing along a stream tube is everywhere constant.

Use of the basic equations permits many fluid flow situations to be analyzed, provided that energy losses are small and can be neglected. In certain special situations, application of the continuity, Bernoulli, and momentum equations permits the energy losses to be approximately computed. *See* HYDRAULIC JUMP; SHOCK WAVE.

For the vast majority of real-fluid-flow situations, experimental information is required to determine the amount of mechanical energy converted to thermal energy. The effect of losses in steady flow of a fluid along a streamline may be expressed by including one or more loss terms in the Bernoulli equation. The equation states that the energy per unit weight at one point is equal to the energy remaining at a downstream point plus all the losses between the two points. For example, in turbulent pipe flow the losses due to wall friction tend to vary almost directly as the length of pipe, inversely as the diameter, and directly as the square of the velocity. Fluid properties and condition of the pipe wall surface enter in a more complicated manner. Experimentally the losses may be determined and expressed in a dimensionless form by a chart so that the results apply to other fluids and to other sizes of geometrically similar pipe. Losses in pipe flow due to changes in cross section or direction, such as elbows and valves, tend to vary about as the square of the velocity, or as a constant [illegible]

dynamically similar if (1) they are geometrically similar, (2) they have the same boundary condi- [illegible]

The particular method of interpreting the data from model studies depends upon the types of forces that predominate (*see* DYNAMIC SIMILARITY). For example, with hydraulic structures such as dams, spillways, and canal transitions, the important forces are those due to gravity and inertia of the liquid, with viscous forces of lesser importance. The ratio of inertial forces to gravity forces is a dimensionless parameter, known as the Froude number. By adjusting flows and depths so that the Froude number is the same in model and prototype, measurements in the model can be made and converted to corresponding prototype values. If the ratio of linear dimensions of the prototype to the model is λ, then the velocity in the prototype is $\sqrt{\lambda}$ times the corresponding velocity in the model, and (for the same liquids in model and prototype) the pressure intensity in the prototype is λ times the corresponding pressure intensity in the model. *See* FROUDE NUMBER.

Model studies of fluid machinery. In making a test on a model of a turbomachine, special relationships must be observed. For geometrically similar streamlines, the ratio of velocity of flow at some point in the machine to the peripheral velocity of the runner or rotor must be the same in model and prototype. Also, since inertial forces are of [illegible]

tion being the same for model and prototype. In addition, for compressible flow, the ratio of veloci[illegible] to local acoustic velocity must be the

at corresponding points in model and prototype. The above conditions, together with geometric similitude, permit tests on a model to be used to predict performance of the prototype. The above relations, referred to as homologous relationships, do not permit viscous forces to be scaled properly, and hence there is a slight difference in efficiency of the various sizes. The larger the machine, the more efficient it is, but with the change being usually not more than 2 or 3%.

In flow through pipes and other closed conduits, the controlling forces are inertial and viscous, for velocities that are small compared with acoustic velocity. The ratio of inertial to viscous forces is expressed by a dimensionless parameter known as Reynolds number (*see* REYNOLDS NUMBER). When two geometrically similar closed flow systems have the same Reynolds number at corresponding points, the dimensionless flow and loss coefficients will be the same. When closed channel flow occurs at velocities near acoustic velocity, the Mach number (ratio of velocity to acoustic velocity) becomes a controlling parameter.

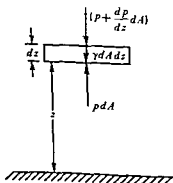
[V.L.S.]

Bibliography: H. Rouse (ed.), *Advanced Mechanics of Fluids*, 1959; V. L. Streeter, *Fluid Mechanics*, 2d ed., 1958.

Fluid statics

The determination of pressure intensities and forces exerted by liquids and gases at rest. Hydrostatics, although implying the statics of water alone, applies to liquids in general (*see* HYDROSTATICS). By definition, a fluid at rest cannot sustain a shear stress, therefore a fluid force exerted on an element of boundary area must act normal to the area. By considering the small free body of fluid shown in the illustration, equilibrium requires that $dp = -\gamma\, dz$, in which p is the absolute pressure intensity, γ is the specific weight (weight of a unit volume of fluid), and z is the elevation, measured vertically upward. For liquids, γ is substantially constant, and the equation shows that pressure intensity decreases linearly as the elevation increases.

With gases, the variation of γ with pressure or elevation must be known in order to integrate the



Free-body diagram for vertical forces acting on fluid element.

equation. For an isothermal gas ($p/\gamma = p_0/\gamma_0 =$ constant) the pressure variation with elevation is given by

$$p = p_0 \exp\left[-(z - z_0)/(p_0/\gamma_0)\right]$$

in which p_0 and γ_0 are the values of p and γ at elevation z_0.

On the average, the absolute temperature T of the atmosphere tends to decrease linearly with elevation within the troposphere, $T = T_0 - \beta z$. By using the general gas law $p/\gamma = RT$, in which R is the gas constant, the specific weight is given by

$$\gamma = p/R(T_0 - \beta z)$$

Use of this expression for γ in terms of p and z yields the variation of pressure intensity with elevation in a standard atmosphere

$$p = p_0(1 - \beta z/T_0)^{1/R\beta}$$

in which p_0 and T_0 are the values at $z = 0$.

The determination of fluid pressure intensity is of importance in many flow measuring devices. The manometer is one method, which makes use of transparent tubes with one or more liquids contained in them. By measuring the difference in elevation of the fluid menisci, the desired pressure intensity is determined from the laws of fluid statics. *See* PRESSURE MEASUREMENT.

Fluid forces on plane surfaces may be determined by integration of $p\,dA$ over the surface, with p the pressure intensity and dA an element of the surface area. For liquids the magnitude of the force is the product of the area and the pressure intensity at the centroid of the area. The line of action of the resultant force is normal to the surface and acts at a point called the center of pressure. For horizontal submerged surfaces the centroid of the area and the center of pressure coincide, but for all other orientations of the surface, the center of pressure is below the centroid; that is, at less elevation.

To determine the resultant fluid force exerted on a curved surface, it is convenient to consider horizontal and vertical components of the force. The horizontal component of force exerted on a curved surface is equal to the force exerted on a projection of the surface onto a plane normal to the direction of the component. The line of action for the horizontal force will be through the center of pressure of the projected area. The vertical component of force exerted on a curved surface by the fluid pressure equals the weight of liquid vertically above the curved surface, and acts through the centroid of this volume of liquid. By taking two horizontal components at right angles and the vertical component, the resultant force is the sum of these three vectors. *See* ARCHIMEDES' PRINCIPLE; BUOYANCY; FLUID MECHANICS.

[V.L.S.]

Fluid-flow principles

Fundamentals of fluid dynamics that govern motion phenomena. Fluid dynamics may be divided into hydrodynamics and gas dynamics. Hydrodynamics

relates to the motion of liquids and gases where density changes are negligible. In general, gases flowing at speeds much less than the speed of sound may be approximated as incompressible. Only the mechanical principles of (1) mass conservation and (2) momentum conservation are necessary to determine hydrodynamic fluid motions. Gas dynamics treats the motion of gases where significant density changes occur, and it requires, in addition to the principles used in hydrodynamics, the principles of thermodynamics, namely, (3) the conservation of energy (first law of thermodynamics), (4) the equation of state of the gas, and in some cases also (5) the no-perpetual-motion-machines law (second law of thermodynamics).

For chemically reacting fluids, flows involving both the liquid and gas phases of a substance, and inhomogeneous mixtures of different fluids, concepts and principles from physical chemistry must be added. For electrically conducting fluids in the presence of electromagnetic fields, electromagnetic body forces occur, requiring the introduction of concepts and principles from electromagnetic theory. In general, only homogeneous nonreacting fluids with zero electrical conductivity are considered here. *See* GAS DYNAMICS; HYDRODYNAMICS; MAGNETOHYDRODYNAMICS; PLASMA PHYSICS; THERMODYNAMIC PRINCIPLES.

Flow continuity. The most common way of describing a fluid-flow field is by giving velocity $\mathbf{v}$, pressure p, density ρ, and temperature T as functions of position in space and time.

Mass. The principle of mass conservation states that the rate of increase of fluid mass in any volume V, fixed in space, equals the net rate of mass flow into this volume. If S is the surface of this volume with $\mathbf{n}$ a unit vector normal to the surface, this principle is stated analytically as

$$\frac{\partial}{\partial t}\iiint_V \rho\, dV = -\iint_S \rho\mathbf{v}\cdot\mathbf{n}\, dS$$

By the use of Gauss' divergence theorem the surface integral can be written as a volume integral

$$\iint_S \rho\mathbf{v}\cdot\mathbf{n}\, dS = \iiint_V \nabla\cdot(\rho\mathbf{v})\, dV$$

which implies that

$$\iiint_V \left[\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{v})\right] dV = 0$$

for any volume V. This in turn implies that the integrand itself must vanish everywhere so that

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) = 0$$

This is called the equation of continuity.

Momentum. The principle of momentum conservation is a statement of Newton's particle laws applied to the many-particle fluid system. The rate of increase of momentum in a fixed volume must equal the net rate at which momentum is convected into the volume plus the forces acting on the volume. The analytical statement for an inviscid fluid is

$$\rho\frac{\partial\mathbf{v}}{\partial t} + \rho(\mathbf{v}\cdot\nabla)\mathbf{v} = -\nabla p + \rho\mathbf{F}$$

where $\mathbf{F}$ is body forces per unit mass (*see* EULER'S MOMENTUM THEOREM; NAVIER-STOKES EQUATION). For irrotational hydrodynamic flows it is necessary only to find a velocity field such that the flow passes around solid boundaries and the pressure is determined post facto by Bernoulli's theorem. *See* BERNOULLI'S THEOREM; *see also* KELVIN'S CIRCULATION THEOREM

Energy conservation. The principle of energy conservation (first law of thermodynamics) brings in the concepts of temperature, heat, and internal energy. Applied to a fixed volume in the fluid flow it states that the rate of increase of internal energy in any fixed volume in space is equal to the net rate of flow of internal energy into the volume, plus the net rate of heat flow into the volume, plus the rate at which work is being done on the volume. In general, it is only near the boundaries of solid objects in the fluid flow or in shock waves that the heat conductivity and viscosity of fluids must be considered, for these are the regions where high temperature and velocity gradients occur. *See* BOUNDARY-LAYER FLOW; SHOCK WAVE.

Elsewhere in the fluid flow, the changes of state of the fluid particles are closely adiabatic and reversible (*see* ISENTROPIC FLOW). If all fluid particles originate from a zone of uniform state, that part of the fluid flow that does not pass through shock waves or enter into boundary layers is isentropic. Many of the flows encountered in gas dynamics are well approximated as isentropic flows. In boundary layers and shock waves entropy production occurs through the processes of viscous dissipation (work done by viscous shear forces) and heat addition by conduction, which are phenomena outside the considerations of classical thermodynamics.

As mentioned previously, considerations of internal energy and heat are, in general, not necessary to determine hydrodynamic flow fields. The reason for this is that the fluid kinetic energy per unit mass is so small compared to the internal energy per unit mass that the internal energy remains

transfer at low speeds, the temperature field can be determined from the velocity field but does not affect the velocity field. Natural convection heat transfer involves buoyant forces due to density changes and yet involves speeds much less than sound speed; hence it is an exception to the earlier statement that low speed velocity fields are deter-

mined only by mass and momentum considerations. *See* HEAT TRANSFER.

The thermal equation of state of a fluid is a relationship among the variables, pressure p, density ρ, and temperature T, which is valid when the fluid is in thermodynamic equilibrium. Similarly, the caloric equation of state relates the internal energy e to p and T. These relationships are derived from experiments or, where possible, statistical mechanics or kinetic theory (*see* KINETIC THEORY OF MATTER). The assumption of local thermodynamic equilibrium in the fluid is usually sufficiently accurate, although changes from one state to another necessarily involve small relaxation times which, in some cases, may not be negligible, for instance, the flow of a gas such as carbon dioxide through a shock wave.

The second law of thermodynamics introduces the concept of entropy and is necessary in some gas dynamic flows to determine the direction in which certain flow changes can or cannot occur (*see* ENTROPY). For example, the flow through a stationary normal shock wave appears to be possible both from subsonic to supersonic and from supersonic to subsonic until the entropy changes are considered; then it is clear that the subsonic to supersonic shock cannot occur spontaneously in nature. The second law is also the basis of the law of mass action, which is essential in determining equilibrium concentrations of various gases in a mixture of gases; a partially dissociated gas is such a mixture (molecular gas and atomic gas), and an electrically neutral partially ionized gas or plasma is a mixture of atomic species, ions, and electrons.

[A.E.BR.]

Fluid-flow properties

The physical properties of fluids that are significant under dynamic conditions. These properties relate directly to the different basic types of fluid motion.

Properties of liquids and gases. The characteristic feature of a fluid (as opposed to a solid) is its *inability* to offer any resistance to change of shape while remaining in equilibrium. Shearing forces *can exist in a fluid* which is in motion and are associated with the physical property of viscosity. *These shearing* forces vanish however when the fluid is at rest. *See* VISCOSITY OF GASES; VISCOSITY OF LIQUIDS; VISCOUS FLOW.

[illegible]

some physical *property such as* the pressure, density, or temperature at a certain point in a region of fluid flow it is *implied that the* property refers to a small particle of the fluid at the point considered. Viewed against a microscopic *scale*, however, the small particle of fluid must be large enough to contain a great number of individual *molecules*.

The physical properties of a gas can be explained in terms of molecular motion by means of the kinetic theory of gases (*see* KINETIC THEORY OF MATTER). Viewed as a fluid-flow problem, however, a gas is usually treated as a continuum when its bulk motion is being considered. There are exceptions to this in certain extreme cases of gas flow at low pressures, where the motion of individual molecules must be taken into account. *See* SUPERAERODYNAMICS.

Gases are compressible in that the density of a gas depends on the pressure. Changes of density in a gas are of special significance in problems of high-speed flow. These compressible-flow properties are reviewed below under the heading gas dynamics. There are many cases of flow of gas or air at relatively low velocities, however, where density changes are not significant and where the flow is essentially incompressible. *See* SUBSONIC FLIGHT.

Liquids, although not absolutely incompressible, undergo only small changes of volume over a wide range of pressure. The flow of liquids, together with that of gases at low velocities, may therefore be grouped under the heading incompressible flow. When considering the flow of liquids it is important to distinguish between those cases where there is a free surface and other cases where the flow is fully submerged. If there is a free surface it is necessary to take account of the effect of gravitational force on the mechanics of the flow. If the flow is fully submerged, on the other hand, the only effect of gravity is to superimpose the normal hydrostatic pressure on the local pressure *distribution* associated with the motion of the fluid (*see* FLUID STATICS). Another physical property associated with the presence of a free surface *on a* liquid is that of surface tension. This has a significant effect in a restricted range of flow problems where thin films of liquid or sprays of small liquid drops are involved. *See* SURFACE TENSION.

Analysis of fluid flow. There are two basically different ways of studying fluid motion. In the *first* method a flow system is considered and attention is directed to a definite region of the flow bounded by a fixed control surface. A simple example of this method is shown for a short section of a *pipeline* (Fig. 1). It is then possible to apply the principles of conservation of mass, momentum, and energy *to* the net flow of fluid crossing the control surface. This method leads to the three basic *equations* of continuity, momentum, and energy, which are commonly employed in the analysis of *flow* through pipes, ducts, nozzles, turbines, pumps, and flow-measuring devices. *See* FLUID-FLOW PRINCIPLES.

In the second method attention is directed to a definite particle of fluid. The velocity and *acceleration* of the particle are analyzed following the motion of the fluid. The various forces acting *on the* particle of fluid (such as normal pressure, shearing stresses due to the action of viscosity, and gravitational force) are expressed in mathematical terms and the resultant force is equated *to the* rate of

change of momentum of the particle using the ordinary principles of mechanics. This method leads to the formulation of the general equations of motion of a fluid. For the case of viscous flow of a gas or liquid the equation of motion is known as the Navier-Stokes equation (*see* NAVIER-STOKES EQUATIONS). The difficulty in making practical use of this method lies *in solving the equations for the* appropriate boundary conditions in an actual problem

The method of dimensional analysis is of importance in comparison of the flow-patterns and relation of the measurements obtained in different cases of *fluid* flow having geometrically similar boundaries (for example, in relating model tests carried out in a wind tunnel to the actual behavior of a full-scale airplane wing in flight). For this purpose it is important to note the physical dimensions of the various properties and variables used

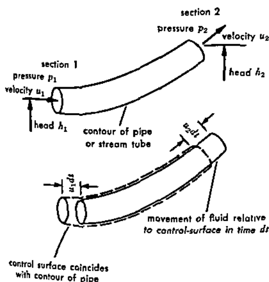


Fig. 1. Example of a flow system for a short length of pipe or stream tube with the simplifying assumption of one-dimensional flow; that is, variation of fluid velocity across each section is ignored.

to describe fluid flow. These are expressed in terms of the basic physical *dimensions of mass, length,* and time (MLT system), or alternatively in terms of force, length, and time (FLT system). The most important quantities used in the study of fluid flow are listed below together with their physical dimensions. *See* DYNAMIC SIMILARITY.

Steady and unsteady flow. *It is important to* distinguish between flow problems in which the motion is steady, that is, where there is no variation of the flow pattern with time, and problems in which the motion is unsteady. In steady flow it is possible to visualize the flow pattern by means of streamlines *which* are drawn so as to coincide at every point with the direction of motion of the fluid at the point considered.

Dimensions of principal flow properties

Quantity	Usual symbol	Dimensions (MLT system)	Dimensions (FLT system)
Length	l	L	L
Fluid velocity	u or v	LT^{-1}	LT^{-1}
Acceleration	du/dt	LT^{-2}	LT^{-2}
Density	ρ	ML^{-3}	FT^2L^{-4}
Force	F	MLT^{-2}	F
Pressure	p	$ML^{-1}T^{-2}$	FL^{-2}
Viscosity	μ	$ML^{-1}T^{-1}$	FTL^{-2}
Kinematic viscosity	ν	L^2T^{-1}	L^2T^{-1}
Shearing stress	τ	$ML^{-1}T^{-2}$	FL^{-2}
Surface tension	σ	MT^{-2}	FL^{-1}
Sonic velocity	a	LT^{-1}	LT^{-1}

If the action of viscosity is ignored, the mathematical analysis of the steady two dimensional flow of an ideal fluid becomes relatively straightforward. This theory of ideal inviscid flow, however, leads to paradoxical results which are at variance with observation of the behavior of real fluids. In particular there would be no drag exerted by an inviscid fluid flowing past an airfoil or streamlined body. Furthermore, regardless of the shape of the body the flow pattern would be of perfect streamline form. *See* D'ALEMBERT'S PARADOX.

It is observed in practice that even with fluids of small viscosity, the exact nature of the flow close to the boundary or surface of the body exerts a profound influence on the flow pattern and on the measured drag. This action of viscosity is discussed below under the heading boundary-layer flow.

Laminar and turbulent flow. *Observation of the* steady flow of a fluid in a pipe or in a boundary-layer region reveals the fact that the character of the motion may be of two entirely different kinds. For example, at relatively low speeds and in pipes of relatively small diameter the fluid is usually observed to flow in a smooth and regular manner. If direct visualization of the flow is achieved by using a glass pipe and by introducing a thin filament of coloring material, the fluid has the appearance of

of small eddies and disturbances in the flow which cause a rapid transverse diffusion of the coloring material This irregular eddying type of motion is known as turbulence The study of turbulent flow

in the case of flow through pipelines, pumps, compressors, and jets, and in the wake behind a body *moving* through a fluid. *See* TURBULENT FLOW.

The first systematic observations on laminar and turbulent flow were made by O. Reynolds. The principal criterion as to whether the flow in a pipe will be laminar or turbulent is the numerical value of the Reynolds number which is defined by

$$Re = \frac{\rho UD}{\mu} \qquad (1)$$

where U is the mean velocity of flow in the pipe, D is the diameter, ρ is the density, and μ is the viscosity of the fluid. If the Reynolds number is less than 2000 approximately, the flow is normally laminar. At higher values of the Reynolds number the flow is usually turbulent. For the case of flow in a boundary-layer region the criteria which determine whether the flow is laminar or turbulent are more complex. *See* BOUNDARY-LAYER FLOW.

For analytical purposes turbulent flow may be regarded as the result of superimposing a random eddying motion, with rapidly fluctuating velocity components of turbulence, on a steady mean flow pattern. It is only within recent years that any progress has been made in developing a satisfactory theory of turbulence. The modern approach to this property of fluid flow is the statistical theory of turbulence which was first proposed by G. I. Taylor.

Boundary-layer flow. It is observed in practice that a real fluid does not slip past a solid boundary. If, for example, a flat plate is inserted in a fluid stream and held in position with the plane of the plate parallel to the streamlines, the fluid immediately adjacent to the plate is brought to rest. If the undisturbed velocity of the fluid is relatively large, it is found that there is a thin layer of fluid adjacent to the plate in which the flow velocity increases from zero (at the surface of the plate) to the undisturbed velocity of the main stream at a short distance from the plate. This region is known as the boundary layer. As in the case of a pipe, the flow in a boundary-layer region may be either laminar or turbulent in character. Typical laminar and turbulent velocity profiles are shown (Fig. 2).

The theory of boundary-layer flow was developed originally by L. Prandtl. The investigation of flow phenomena in boundary layers is of special importance because the boundary-layer region is the origin of the skin friction or drag force exerted by the fluid when flowing past a solid body or wall. Furthermore, the nature of the flow outside the boundary layer may be drastically affected by the flow properties within the boundary layer itself. For example, when there is an adverse pressure gradient, as in the diffuser section of a venturi tube where the pressure is increasing in the direction of the flow, the phenomenon of flow separation or break-away from the wall is determined by the nature of the flow in the boundary layer.

In the case of a laminar boundary layer the shearing stress at any distance y from the wall is related to the transverse velocity gradient by the expression

$$\tau = \mu(du/dy) \tag{2}$$

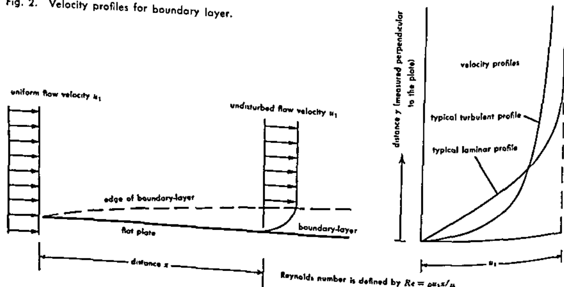
where μ is the coefficient of viscosity of the fluid. A physical explanation of viscosity is provided by kinetic theory. Viscous stresses are set up wherever there is a velocity gradient in a fluid due to the transfer of momentum by the random thermal motion of the molecules of the fluid. In the case of most fluids the viscous stresses are directly proportional to the velocity gradients. *See* NEWTONIAN FLUID.

The skin friction force, or shearing stress τ_0 at the wall, may be expressed conveniently by defining a dimensionless skin friction coefficient $c_f = 2\tau_0/\rho U_1^2$. An exact solution of the equations of motion for laminar flow in a boundary layer may be obtained for the case of a flat plate. This solution yields the following relationship between the skin friction coefficient and the Reynolds number, where D in Eq (1) is replaced by a distance x measured from the leading edge

$$c_f = 0.664Re^{-1/2} \tag{3}$$

This result applies only to a laminar boundary layer and is a special case of the general functional relationship between the skin friction coefficient and the Reynolds number.

Fig. 2. Velocity profiles for boundary layer.



At large values of the Reynolds number the flow pattern in a boundary layer is usually turbulent. Transition from laminar to turbulent flow in the boundary layer may result from various causes: the action of an adverse pressure gradient, roughness of the surface, or the amplification of small disturbances due to instability over a certain range of flow conditions. There is no exact mathematical solution for turbulent flow in a boundary layer. The shear stresses in turbulent flow are greater than in laminar flow owing to the operation of a different momentum-transfer process due to the presence of transverse eddies associated with the turbulent motion. These additional shearing stresses caused by the turbulent velocity components are known as the Reynolds stresses. Although it is not possible to determine an exact theoretical expression for the skin friction coefficient in turbulent flow, measured results may be correlated by the same method of dimensional analysis and the theory of dynamic similarity. For example, in the case of turbulent boundary-layer flow past a flat plate having a smooth surface, the skin friction coefficient will still be a function only of the Reynolds number and results may be expressed graphically in the form of an experimental curve relating c_f and Re. Over a limited range of Reynolds number for turbulent flow in a boundary layer this experimental relationship may be expressed approximately by $c_f = 0.059Re^{-1/5}$.

There is some evidence that, even with fully developed turbulent flow in the main part of the boundary layer, the turbulent velocity fluctuations or eddies must die out as the surface of the plate or wall is approached. The final mechanism by which the shearing stress is transmitted to the wall (as an effective skin friction force) is by the viscous shearing of a thin layer of fluid. This region of the fluid is known as the laminar sublayer.

Fluid flow in pipes. A friction factor or skin friction coefficient for fluid flow in a pipe may be defined in a similar manner to that for a boundary layer. The fluid velocity will be a maximum at the center of the pipe, but because the exact form of the velocity distribution is not usually known, it is convenient to work in terms of the mean velocity of flow in the pipe U (averaged over the cross section). Defining the friction factor f by

$$f = 2\tau_0 / \rho U^2 \qquad (4)$$

where τ_0 is the shear stress at the wall of the pipe, leads to a general expression for the pressure drop $(p_1 - p_2)$ over a length L of a pipe having diameter D and with fluid of density ρ flowing with mean velocity U as

$$(p_1 - p_2) = 2f(L/D)\rho U^2 \qquad (5)$$

This is known as the Fanning equation for pipe friction. See PIPE FLOW.

As in the case of boundary-layer flow, the motion of a fluid in a pipe may be either laminar or turbulent. The Reynolds number for flow in a pipe is defined in terms of mean velocity U and pipe diameter D. For small values of the Reynolds number (up to 2000 approximately) the flow in a pipe will usually be laminar and the velocity profile for fully developed laminar flow (at some distance from the entry to the pipe) will be parabolic in form. The friction factor for laminar flow in a pipe is related to the Reynolds number by $f = 16/Re$ and the pressure drop in the pipe is proportional to the velocity of flow. This is known as the Poiseuille law.

At higher values of the Reynolds number, flow becomes turbulent and, as in the case of turbulent boundary-layer flow, recourse must be had to semiempirical expressions for the friction factor and for the form of the velocity profile. For smooth pipes at very high values of the Reynolds number the following logarithmic expression may be used for the friction factor:

$$1/\sqrt{f} = 4 \log_{10} (Re\sqrt{f}) - 0.40 \qquad (6)$$

and the velocity distribution can best be described by means of the universal velocity profile. The roughness of the wall of the pipe will affect both the friction factor and the exact form of the velocity profile. For practical design purposes it is usually necessary to use measured values of the friction factor obtained from experimental observations of pressure drop and flow in commercial pipes.

Dynamic similarity. The measurement of the drag exerted by a fluid flowing past a streamlined body is a matter of considerable practical importance. Model tests are frequently used for this purpose, for example, wind-tunnel tests on aircraft models and water-channel or tank tests on models of ship hulls (see TOWING TANK; WATER TUNNEL; WIND TUNNEL). The question arises as to the conditions under which the flow pattern in the model test will be similar to the flow pattern for the full-scale aircraft or ship.

The first requirement is that the boundaries of the flow in the two cases should be geometrically similar. This means not only that the model should have exactly the same shape as its full-scale counterpart, but also that the walls of the tunnel or tank used for the test should be at a sufficient distance from the model so as not to influence the flow pattern around the model itself. Taking the relatively simple case of flow past a streamlined section or body, at moderate speeds where both viscous and inertia forces are significant but where compressibility effects may be ignored, the flow

Reynolds number implying large inertia forces and a low value implying that the viscous forces are relatively large. If the Reynolds number is the same for the model test as in the full-scale flow, the flow patterns will be dynamically similar. If the force or drag on the body is measured and exp as a dimensionless drag coefficient,

$$c_D = \frac{2F}{\rho U^2 L^2} \quad (7)$$

where ρ is the fluid density, U is the undisturbed velocity of the stream, and L is a characteristic length, the tests on the model, carried out over an appropriate range of flow conditions, may be correlated by plotting an experimental curve of the drag coefficient against the value of the Reynolds number. For geometrically similar models, in other words, drag coefficient is a function only of Reynolds number.

If the velocity of flow is sufficiently large for effects of compressibility to be important, however, the Reynolds number by itself is not a sufficient criterion of dynamic similarity. In the case of high-speed flow, involving flow velocities of the same order as the velocity of sound in the fluid, it is necessary that the Mach number should have the same value in the model test as in the full-scale flow. *See* COMPRESSIBLE FLOW; MACH NUMBER.

If there is a free surface, as in the case of flow of a liquid, it is also necessary to take account of gravitational forces. The total resistance of a ship, for example, is made up of skin friction drag and also wave-making drag. For the experimental study of the wave-making resistance it is necessary to specify equality of the Froude number between the model test and the full-scale ship. The Froude number, which is another dimensionless quantity, gives a measure of the ratio of inertia to gravity forces. *See* FROUDE NUMBER.

Flow past a streamlined body. The drag exerted by a fluid flowing past a submerged body arises from two causes. First is the skin friction drag, which could be computed from a knowledge of the tangential or shearing stresses exerted by the fluid at every point of the surface. Second is the form drag which is contributed by the distribution of the normal pressure of the fluid over the surface of the body, that is, the stress component acting at right angles to the surface of the body at each point. A bluff-shaped section such as a cylinder has a relatively large form drag because the normal pressure distribution over the rear half, or downstream side, is less than that exerted over the front half of the section (Fig. 3). This is caused largely by the phenomenon of flow separation or breakaway and is associated with the formation of a turbulent or eddying wake. In the case of two-dimensional flow past a bluff obstacle, vortices may be formed at the points of separation and these will travel downstream in the wake, forming a vortex street. Under other conditions the eddying pattern in the wake may be irregular and the wake flow may be completely turbulent. *See* KARMAN VORTEX STREET; WAKE FLOW.

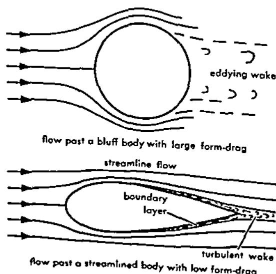


Fig. 3 Flow past a streamlined body.

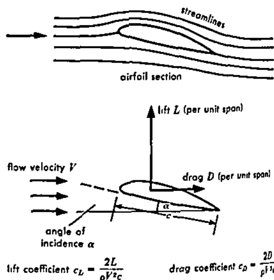


Fig. 4 Flow past an airfoil.

The purpose of streamlining is to reduce the form drag to a minimum. The skin friction drag however, cannot be eliminated. It arises from the property of viscosity of the fluid in the same manner as the skin friction force or shearing stress in boundary-layer flow past a flat plate. *See* STREAMLINING.

Flow past an airfoil. An airfoil section is designed to give a relatively large lift force (measured at right angles to the direction of flow) and as small a drag force as possible. The lift and drag forces may be expressed as dimensionless coefficients and the usual notation is shown on the accompanying sketch (Fig. 4). The lift coefficient will depend on the angle of incidence measured between the chord of the airfoil and the direction of flow. The lift coefficient increases with the incidence up to a certain maximum value which will depend on the design of the particular airfoil section. At higher values of the incidence, the lift coefficient falls sharply because of the separation of the flow over the upper surface. This phenomenon is known as stalling. Airfoil theory is of practical importance, not only in the design of aircraft, but also in connection with turbines and compressors

where the blading may be of airfoil section. *See* AIRFOIL.

Gas dynamics. Compressibility effects become significant in the flow of a gas if the velocity is comparable to the sonic velocity. At subsonic velocities the main effect is simply the variation in the density of the gas and the consequent introduction of an additional variable into the equations of motion. At supersonic velocities, however, the flow changes in character. Discontinuities can occur in the pressure, density, and velocity of the fluid. These finite disturbances are known as shock waves. *See* SHOCK WAVE.

Sonic velocity is simply the speed of propagation of an infinitesimally small disturbance or sound wave in the fluid. Finite disturbances, however, can travel at greater velocities. If a thin sharp-edged body is moving relative to the fluid at a supersonic velocity, the pressure disturbance produced by the body cannot be propagated throughout the fluid. In the case of a sharp conical-headed projectile, for example, a shock wave is formed (Fig. 5), and the air on the upstream side ahead of the shock wave receives no advance indication of the presence of the projectile. In the case of a bluff-nosed body moving at supersonic speed a curved detached shock wave is formed and moves with the body at a short distance ahead of the nose. *See* SUPERSONIC FLIGHT.

Flow through a shock wave is associated with a sudden increase in pressure and density and a reduction in velocity. The simplest case is that of a normal shock wave, which is a plane shock front at right angles to the direction of motion. Viewing the shock front as stationary, flow through a normal shock wave always involves a change from supersonic conditions on the upstream side to subsonic flow of the gas on the downstream side. Oblique shock waves are associated with a change in the direction of flow of the gas. The flow on the upstream side of an oblique shock wave must be supersonic, but that on the downstream side may be either subsonic or supersonic, depending on the inclination of the wave to the direction of flow of the incident stream. *See* COMPRESSIBLE FLOW; SHOCK-WAVE DISPLAY; SUPERSONIC DIFFUSER.

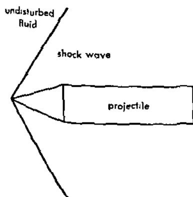


Fig. 5. Supersonic flow past a projectile

In addition to shock waves, which are waves of compression, it is possible for expansion waves to be formed in supersonic flow. An important example is that of the Prandtl-Meyer expansion where a supersonic stream of gas expands in flowing round a corner or convex boundary. The flow past a supersonic airfoil is always characterized by a pattern of oblique shock waves and expansion waves. *See* PRANDTL-MEYER EXPANSION FAN.

Forced convection. In many engineering applications, the heat-transfer properties of fluid flow are important, for example, in the design of boilers, heat exchangers, nuclear reactors, chemical plants, and rockets. Similarly in the field of chemical engineering, the mass-transfer properties of fluid flow are of equal importance. Heat or mass transfer between the wall of a pipe, or any other surface or boundary of the flow, and a fluid flowing past the wall is known as forced convection if the motion of the fluid is determined by external forces. *See* CONVECTION (HEAT).

The mechanism of heat and mass transfer by forced convection is essentially similar to that of momentum transfer due to the action of viscosity and turbulence which causes the effect of skin friction. There is thus an analogy between skin friction, heat transfer, and mass transfer in fluid flow. Heat transfer by forced convection in turbulent flow is the type most frequently encountered in engineering applications. [J.M.KA.]

Bibliography: G K. Batchelor (ed.), *The Scientific Papers of Sir Geoffrey Ingram Taylor*, 1959; S. Goldstein (ed.), *Modern Developments in Fluid Dynamics*, 1938; L. Howarth (ed.), *Modern Developments in Fluid Dynamics, High Speed Flow*, 1953; J. C. Hunsaker and B G. Rightmire, *Engineering Applications of Fluid Mechanics*, 1947; C. Jaeger, *Engineering Fluid Mechanics*, 1957; J. M. Kay, *An Introduction to Fluid Mechanics and Heat Transfer*, 1957; H. W. Liepmann and A. E. Puckett, *Introduction to Aerodynamics of a Compressible Fluid*, 1947; W. H. McAdams, *Heat Transmission*, 3d ed., 1954; L Prandtl, *The Essentials of Fluid Dynamics*, 1952; H. Rouse, *Fluid Mechanics for Hydraulic Engineers*, 1938.

Fluidization of solids

The processing technique employing a suspension or fluidization of small solid particles in a vertically rising stream of gas so that gas and solid come into intimate contact This is a tool with many applications in the petroleum and chemical process industries. There are major possibilities of even greater growth and utility in the future. Suspensions of solid particles by vertically rising liquid streams are of lesser interest in modern

CATION, MECHANICAL.

Properties of fluidized systems. The interrelations of hydromechanics, heat transfer, and mass

transfer in the gas-fluidized bed involve a very large number of factors needed for mathematical and physical analysis and process design. In the usual case, the practical applications in multi-million-dollar plants have far outrun the exact understanding of the physical, and often chemical, interplay of variables within the minute ranges of each of the small particles and the surrounding gas phase.

The fluidized bed, as it is called, results when a fluid, usually a gas, flows upward through a bed of suitably sized, solid particles at a velocity sufficiently high to buoy the particles, to overcome the influence of gravity, and to impart to them an appearance of great turbulence, similar to that of a violently boiling liquid. Fluid velocities must be intermediate between that which would lift the particles to maintain a uniform suspension and that which would sweep the particles out of the container. The fluidized bed is in a relatively stable condition of vigorous contacting of fluid and solids, with a lower boundary at the point of fluid inlet and an upper definite and clearly marked boundary surface at which the gas disengages itself from the system.

Particle sizes are often in the range of 30–125 microns and superficial gas velocities may be in the range of 0.02–1.0 ft/sec, depending on factors such as the relative densities of the gas and solid, the size and shape of solid particles, and the number of particles per cubic foot (bed density).

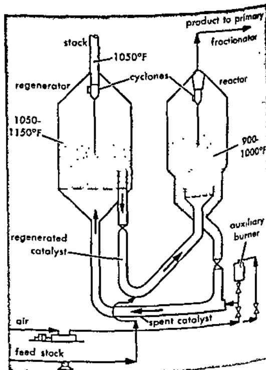
Conditions within the bed are intermediate between that of a packed column and that of a pneumatic transporter of the particles. In pneumatic transport, the slippage of the gas past the solid particles is often relatively small, and the components of friction and inertia carry the solid particle along, whereas in the fluidized bed, the slippage of gas past the solid particle is sufficient to balance the gravitational effect or weight, but is not sufficient to supply a continuing velocity.

An increase in fluid velocity increases the upward drag on the particles. This causes them to lift, thereby increasing the bed voidage and the effective cross section available for gas passage. This decreases the interstitial velocity to reduce the drag on the individual particles, which then settle somewhat until the forces on the particles ultimately come into balance. Further increase in velocity causes further bed expansion until the bed passes beyond the limits of the containing vessel and the particles are transported. There is excellent contact between the solid particles and the gas; and if there is a temperature difference, there is an excellent flow of heat between the particles themselves and between the particles and the gas. Thus, equalization of temperature is rapidly accomplished. Similarly, there is a major opportunity for mass transfer to or from the solid particles and between the solid particles and the gas phase, as in the movement of reactants in the gaseous phase toward the surface of solid particulate catalysts and the reverse movement of products away from the catalytic surface, back to the body of the gas phase.

Principal applications. Thus, with such excellent opportunities for heat and mass transfer to or from solids and fluids, fluidization has become a major tool in such fields as drying, roasting, and other processes involving chemical decomposition of solid particles by heat. An important application has been in the catalysis of gas reactions, wherein the excellent opportunity of heat transfer and mass transfer between the catalytic surface and the gas stream gives performance unequaled by any other system. In the petroleum industry, for example, the cracking of hydrocarbon vapors in the presence of solid catalysts affords many advantages. Among these are a minimum tendency for overheating of one particle or one part of the catalytic bed as compared to another, and the immediate removal of the reactant products, resulting in greatly increased efficiency of the catalyst. The catalyst itself often becomes carbonized on the surface because of the coking tendency of the reacting hydrocarbons. This catalyst may then be removed by pneumatic transport or otherwise to a second reactor wherein an oxidation is accomplished, again with maintenance of relatively uniform conditions. The rejuvenated catalyst may be returned to the prime reactor by pneumatic or other transfer lines, and the process can be continued indefinitely.

Fluid reaction systems have been standardized in design, permitting the operation of units comparable in maximum size to other types of units.

From a process development standpoint, the dense bed of particles and gas means that enough catalyst can be maintained in a relatively small vessel to give a high degree of conversion with a low



Flow diagram of a catalytic cracking process using fluidized catalyst.

pressure drop. The mechanical design of a fluidization system and accessories is comparable in many respects to that of a body of liquid. However, special devices are needed to control fluid and solid handling, to introduce gas to the fluid bed, and to control the level of the solid. Vapor disengaging units, filters and cyclones for removal of minute amounts of fines which might otherwise escape, blowers, pumps, pneumatic transport lines, and bucket elevators are also needed.

One important component in many units, besides the fluidizing chamber—often the reactor—is the standpipe, which allows a recycling of solids and builds up pressure on the solids for transfer purposes. In a large catalytic cracking unit, the rate of circulation of solids between vessels may amount to as much as 50 tons/min. See BULK-HANDLING MACHINES; CATALYSIS; CRACKING; UNIT OPERATIONS. [D.F.O.]

Bibliography: D. F. Othmer (ed.), *Fluidization*, 1956; F. A. Zenz, *Fluid catalyst design data*, *Petroleum Refiner*, vol. 36, nos. 4–8, 1957; F. A. Zenz and D. F. Othmer, *Fluidization and Fluid-Solid Systems*, 1960.

Fluoborate

In the broadest sense, the term fluoborate refers to a group of compounds related to the borates in which one or more oxygens have been replaced by fluorines. In a strict sense, fluoborate refers to the BF_4^- ion which is derived from tetrafluoroboric acid, HBF_4. Many other fluoborates are known only in the form of salts. Some of the postulated parent acids are H_2BF_5, $H_2B_2F_6$, $H_4B_4F_{10}$, H_2BOF_3, H_2BO_2F, and H_3BO_2F. The lithium, sodium, ammonium, alkaline-earth, and heavy-metal salts of these acids are soluble. The fluoborates are used as fluxes and in electroplating. See BORATE; BORON; FLUORIDE; FLUORINE. [E.E.WR.]

Fluorescence

Fluorescence is generally defined as a luminescence emission that is caused by the flow of some form of energy into the emitting body, this emission ceasing abruptly when the exciting energy is shut off. In attempts to make this definition more meaningful it is often stated, somewhat arbitrarily, that the decay time, or afterglow, of the emission must be of the order of the natural lifetime for allowed radiative transitions in an atom, which is about 10^{-8} sec for transitions involving visible light. Perhaps a better distinction between fluorescence and its counterpart, phosphorescence, rests not on the magnitude of the decay time per se, but rather on the criterion that the fluorescence decay is temperature-independent. If this latter criterion is adopted, the luminescence emission from such materials as the uranyl compounds and rare-earth-activated solids would be called "slow fluorescence" rather than phosphorescence. The decay of their luminescence takes place in milliseconds, rather than in 10^{-8} sec; but the decay is temperature-independent over a considerable range of temperature, and it follows an exponential decay law

$$I = I_0 \exp(-t/\tau_l)$$

that is to be expected for spontaneous transitions of electrons from an excited state of an atom to the ground state when the atom has a transition probability per unit time $1/\tau_l$. In this equation I is the luminescence intensity at a time t, and I_0 is the intensity when $t = 0$. At higher temperatures, where the luminescence efficiency drops because other transitions that can dissipate the energy begin, a temperature-dependence of the decay time can be observed because $1/\tau_{obs} = 1/\tau_l + 1/\tau_{diss}$, where $1/\tau_{diss}$ is the transition probability per unit time for the nonradiative transition. This reservation must therefore be made in accepting the preceding criterion for fluorescence.

The decay time of fluorescent materials varies widely, from the order of 5×10^{-9} sec for many organic crystalline materials up to 2 sec for the europium-activated strontium silicate phosphor. *Fluorescent materials with decay times between* 10^{-9} and 10^{-7} sec are used to detect and measure high-energy radiations, such as x-rays and gamma rays, and high-energy particles such as alpha particles, beta particles and neutrons. These agents produce light flashes (scintillations) in certain crystalline solids, in solutions of many polynuclear aromatic hydrocarbons, or in plastics impregnated with these hydrocarbons.

The so-called fluorescent lamps employ the luminescence of gases and solids in combination to produce visible light. A fluorescent lamp consists of a glass tube filled with a low-pressure mixture of argon gas and mercury vapor, coated on the inside surface with a luminescent powder or blend of such powders, and having an electrode at each end. An electrical discharge is passed through the gas between the two electrodes, exciting the mercury atoms to luminesce. This results in the emission of both visible and ultraviolet light. At the low pressures employed, approximately half of the electrical energy input to the lamp is converted into the 2537 A radiation characteristic of the mercury atom. *The phosphor coating is chosen for the* efficiency with which it is excited by this wavelength of ultraviolet and for the color of visible luminescence that is desired. The small amount of visible light generated by the discharge itself is largely transmitted by the phosphor coating and adds slightly to the luminous output of the lamp.

tion of the electrical energy supplied to them. See FLUORESCENT LAMP; LUMINESCENCE. See also ABSORPTION (ELECTROMAGNETIC RADIATION).

[C.C.K.; J.H.S.]

Fluorescent lamp

A lamp which produces light largely by conversion

phors produce most of the light provided by fluorescent lamps manufactured today.

The fluorescent lamp consists of a glass tube containing two electrodes, a coating of powdered phosphor, and small amounts of mercury. The glass tube seals the inner parts of the lamps from the atmosphere The electrodes provide a source of free electrons to initiate the arc and are connected to the external circuit through the ends of the lamp. The phosphor is a chemical or mixture of chemicals that convert short-wave ultraviolet energy into light (*see* FLUORESCENCE). The mercury, when vaporized in the arc, produces the ultraviolet radiation that causes fluorescence. Argon gas, introduced in small quantities, provides the ions that facilitate starting of the lamp (Fig 1).

Operation. To start the flow of current that forms the arc in the tube, free electrons must be introduced into the tube. This is accomplished by thermionic emission, field emission, or a combination of these techniques of obtaining electron emission from a cathode.

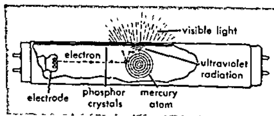


Fig. 1. Fluorescent lamp. (*Westinghouse Lighting Handbook*)

Current passed through the electrodes of a preheat lamp causes emission of electrons from the electrodes into the tube (*see* THERMIONIC EMISSION). When sufficient electrons are released, the resistance of the gap between the electrodes is low enough to permit striking the arc across the gap.

The application of a high potential difference across opposite instant-start cathodes in a lamp draws electrons from the negative electrode (*see* FIELD EMISSION) and attracts them toward the positive electrode. On alternating-current circuits, the electrodes are alternately negative and positive in each half-cycle, so both electrodes emit electrons. Instant-start cathodes may be either hot or cold. The hot cathode permits greater lamp current and produces lower over-all lighting costs; it consists of a coiled wire coated with a material that yields electrons freely. The cold cathode is well suited to sign tubing and lighting lamps of special lengths or shapes; it consists of a thimble-shaped iron cup.

Electrodes that are specially designed to heat quickly can be used in conjunction with a moderately high voltage to attract electrons into the tube. This technique is used with rapid-start electrodes. Rapid-start electrodes are continuously heated during lamp operation.

Basic electric circuits for operation of the three types of fluorescent lamps include the preheat, instant-start, and rapid-start circuits. Each includes a ballast, which limits starting voltage and current (*see* VAPOR LAMP). Many circuit variations are used for the operation of one or more lamps from a single ballast, with and without circuit elements for the correction of power factor.

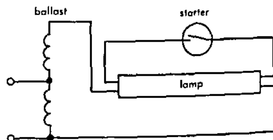


Fig. 2. Preheat lamp circuit.

Preheat circuit. In the preheat circuit (Fig. 2), a starting switch, usually an automatic starter, is used to heat the electrodes. The most common starter employs a small argon glow tube (*see* NEON GLOW LAMP) with one fixed electrode and one electrode made of a bent bimetallic strip. When the circuit is energized, a low-current glow discharge forms in the starter; the heat from this glow is sufficient to expand the bimetallic electrode until it contacts the fixed electrode, forming a short circuit through the starter. The full output voltage of the ballast then causes current to flow through the lamp electrodes, heating them and causing them to emit electrons. The starter cools because the glow is no longer present, and the switch opens, impressing full ballast voltage between the lamp electrodes. If there are enough electrons in the tube, the arc is formed; if not, the glow-switch process is repeated. This process may require several repetitions, accounting for the delay and flickering present at the starting of preheat lamps. When the arc has formed and full lamp current is flowing, the ballast absorbs about half its initial voltage, and there is not sufficient voltage to cause the glow switch to operate; it then becomes an inactive circuit element.

Preheat fluorescent lamps are commercially available in lengths from 6 in. to 5 ft, with wattages from 4 to 100. Longer lamps, or lamps of higher wattage generally use instant-start or rapid-start circuits, which have more favorable starting characteristics. Each size of fluorescent lamp requires a ballast designed to match the requirements of the lamp in starting voltage and current, and operating current.

Instant-start circuit. In the instant-start circuit of Fig. 3 the ballast voltage is much higher than in the preheat circuit. Immediately upon energizing the circuit this voltage attracts enough electrons into the tube to form the arc. For a given lamp wattage and current, the ballast must absorb a higher voltage than with preheat ballasts; hence the ballast is usually larger and dissipates greater wattage. Lampholders for instant-start circuits are connected so that the ballast is disconnected when the lamp is removed, automatically eliminating the

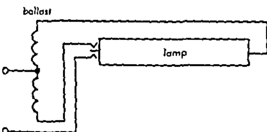


Fig 3. Instant-start lamp circuit.

hazard of high voltage during maintenance operations.

Instant start lamps for general lighting purposes are available in lengths from 42 in. to 8 ft. with wattages from about 15 to 74. Shorter lamps are not usually economical for general lighting service, because they require larger and more costly ballasts than preheat circuits. Higher wattages would require overly large ballasts and cathodes of costly construction.

Rapid-start circuit. The rapid-start ballast contains transformer windings that continuously provide the required voltage and current for electrode heating. When the circuit is energized, these windings quickly heat the electrodes, releasing enough electrons into the tube for the lamp to arc from the voltage of the main windings. This combination of heat and moderately high voltage permits quick lamp starting with smaller ballasts than those for instant start lamps, eliminates the annoying flicker associated with the starting of preheat lamps, and eliminates the starter and its socket from the lighting system (Fig. 4).

The continuously heated cathode of the rapid-start lamp is better adapted to higher lamp currents and wattages. Therefore, rapid-start lamps are available in wattages up to 215 for 8-ft lamps The smallest conventional lighting lamps are 3-ft, 30-watt lamps, but circular rapid start lamps are available in 22-, 32-, and 40-watt sizes.

Special ballasts and circuits are available that permit the economical dimming and flashing of rapid start lamps, providing a range of applications and control that was not possible with previous fluorescent lamp types. Rapid-start lamps are commonly used in flashing signs, and in residential and commercial lighting where continuously variable illumination levels are desired.

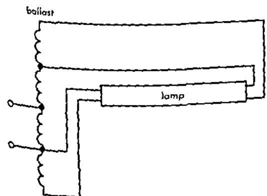


Fig 4 Rapid-start lamp circuit.

High-frequency fluorescent lighting. Fluorescent lamps are usually operated on alternating-current circuits with a frequency of 60 cycles per second (cps). However, higher frequencies permit higher efficiency lamp operation with simpler ballasts of lower power dissipation per watt of lamp. Consequently, systems have been developed for the operation of fluorescent lamps at frequencies from 360 to 1500 cps. These systems employ various types of frequency converters to obtain the higher-frequency power. Motor-generator sets are most common. but static magnetic converters and converters using transistors and controlled rectifiers have also been developed. High-frequency systems can often provide lower over-all lighting cost for fluorescent lamp loads of 25 to 50 kilowatts or more.

Lamp colors. Fluorescent lamps provide light at several times the efficiency of incandescent lamps, the exact ratio depending on the fluorescent lamp color. Lamp color is determined by the selection of chemicals used in the phosphors; various chemicals respond to the ultraviolet energy in the arc by producing different colors of light. Several types of essentially white fluorescent lamps are available commercially, as well as a range of tinted and saturated colors.

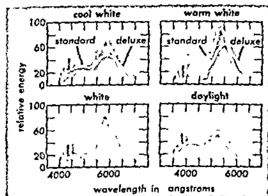


Fig 5. Spectral energy distribution for fluorescent lamps The discontinuities in the curves are characteristic lines of the mercury emission spectrum. (Westinghouse Lighting Handbook)

One of the earliest fluorescent "whites" was the so-called daylight lamp, which produced a bluish-white light of poor color rendition. The most widely-used fluorescent color is called cool white; its lighted appearance is whiter than that of the daylight lamp, it produces more light per watt, and its color rendition is better. In commercial and residential lighting, where the faithful rendition of colors is important, deluxe cool white and deluxe warm white lamps are commonly used; providing vastly superior color rendition at a slight sacri in lamp efficiency. Figure 5 shows the relati tral energy for lamps of equal wattage.

Fluoride

A compound derived from hydrofluoric acid, HF, in which the fluorine atom is in the 1− oxidation state.

The fluorides are different in many respects from the other halides. *See* CHLORIDE; HALIDE.

The small size of the fluoride ion means that the extra electron is held tightly; it is difficult to convert the fluoride ion to free fluorine. The small size also means that there is less tendency to form covalent bonds in simple salts such as mercuric fluoride, HgF_2, which has an ionic lattice whereas its chlorine analog, mercuric chloride, $HgCl_2$, has a molecular crystal. In direct contrast to this, many fluorides are known in which the central atom is in a higher oxidation state than that observed for other halides and these compounds are often covalent, for example, vanadium pentafluoride, VF_5, chromium tetrafluoride, CrF_4, bismuth pentafluoride, BiF_5, and sulfur hexafluoride, SF_6.

A group of particular interest is the hexafluorides. They are very volatile for their relatively high molecular weight of about 300 g/mole. Hexafluorides of sulfur, selenium, tellurium, molybdenum, tungsten, rhenium, osmium, iridium, platinum, uranium, neptunium, and plutonium are known. Platinum hexafluoride, PtF_6, has been prepared only recently.

The fluoride ion is involved in many stable coordination complex ions such as fluoborate, BF_4^-, fluoaluminate, AlF_6^{3-}, fluosilicate, SiF_6^{--}, hexafluophosphate, PF_6^-, and hexafluoroferrate, FeF_6^{3-}. Fluorine also replaces oxygen in anions to give fluorophosphates, fluoborates, fluosilicates, and fluosulfonates. Certain esters of fluophosphoric acid are known as nerve gases, and others have been developed as effective insecticides.

The solubilities of the inorganic fluorides are somewhat different from those of other halides; notably, silver fluoride, AgF, thallous fluoride, TlF, and mercurous fluoride, Hg_2F_2, are much more soluble whereas magnesium fluoride, MgF_2, calcium fluoride, CaF_2, strontium fluoride, SiF_2, barium fluoride, BaF_2, and lithium fluoride, LiF, are much less soluble than the corresponding chlorides.

Of the naturally occurring fluorine compounds, cryolite, Na_3AlF_6, is important in aluminum metallurgy and fluorspar, CaF_2, is used as a flux in metallurgy.

Sodium fluoride is used as an insecticide and rodenticide. Since a small amount of F^- ion (0.8–1.6 ppm) helps to prevent cavities in teeth, the fluoridation of municipal water supplies has become widespread.

Fluoride may be detected analytically by the addition of sulfuric acid. The hydrogen fluoride evolved will etch glass. *See* FLUOBORATE; FLUOROCARBON; FLUOSILICATE; HALOGENATED HYDROCARBON; INSECTICIDE; ORGANOPHOSPHORUS COMPOUND; POLYFLUOROOLEFIN RESIN. [E.E.WR.]

Fluorimetric analysis

A method of chemical analysis in which a sample, exposed to radiation of one wavelength, absorbs this radiation and reemits radiation of the same or longer wavelength. If this reemission occurs in about 10^{-9} sec, it is called fluorescence. Reemission after about 10^{-6} sec or more is called phosphorescence. Fluorescence analysis utilizes the reemitted light to determine the material which is the source of the fluorescence.

The radiation source is usually a mercury arc, and one line of its spectrum is isolated by the primary filter. The light then passes through the sample, and the fluorescence is measured at an angle, usually a right angle, to the light beam. The secondary filter before the detector eliminates scattered radiation of the wavelength of the source. The arrangement of a fluorimeter is very similar to that of a nephelometer, and it has the same advantage of sensitivity—being able to detect very small amounts of reemitted radiation. Fluorimetric analysis can also be employed in the x-ray region using a molybdenum or tungsten source, appropriate metal filters, and, usually, a solid sample.

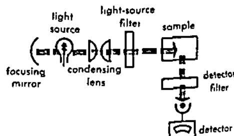


Diagram of a simple fluorimeter.

For quantitative work at very low fluorescent intensities, the intensity is almost directly proportional to the concentration of the fluorescing material. At high fluorescent intensities there are usually appreciable deviations from any linear function of fluorescent intensity and concentration. For this reason, fluorescence analysis is usually carried out in a very dilute solution, and a calibration curve of fluorescent intensity versus concentration of emitter is carefully prepared.

Among the many variables which must be closely watched in quantitative fluorescence analysis are scattering of the incident light by colloidal particles in solution, absorption of the fluorescent light by colored materials in solution, either quenching or intensification of the fluorescence of the compound in question by other ions or compounds present in the solution, fluorescence of other compounds present in the solution, and often the extreme dependence of fluorescence intensity on the temperature of the solution.

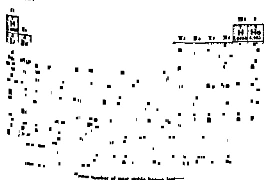
In spite of the drawbacks mentioned above, fluorescence is widely used in analytical work because of its sensitivity and selectivity in many systems.

For example, vitamin B_2 (riboflavin) is usually determined by its fluorescence in solutions as dilute as 0.001 μg/ml. Other compounds or ions commonly determined by fluorescence in solution are β-phenylnaphthylamine, an organic antioxidant; thiamin (vitamin B_1), after reaction with ferricyanide; aluminum, after complexing by 8-hydroxyquinoline; and zirconium, after reaction with morin (an organic compound). Solids are also often analyzed by fluorescence. In particular, uranium compounds fluoresce with a yellow color, and uranium in ores is often detected after fusion of the ore with sodium fluoride. Many sensitive qualitative tests for the presence of various inorganic ions have been developed, based on the selective formation of a fluorescing compound.

Fluorescence spectroscopy has recently been developed to increase the selectivity of fluorimetry. In this technique the emitted fluorescent light is passed through a monochromator so that the fluorescence emission spectrum may be recorded. It is possible to measure several fluorescing compounds in the same solution, provided that they have sufficiently different fluorescence emission spectra. In obtaining this selectivity, some sensitivity is lost, since some fluorescent light is lost in the monochromator, and only a small portion of the total fluorescent energy emitted is measured at any one wavelength. A further increase in selectivity is effected by using different wavelengths of incident radiation to excite the fluorescence. It may be expected that this general technique will be used more in the future since fluorescence spectroscopy eliminates or minimizes many of the errors in fluorescence analysis which were mentioned previously. See FLUORESCENCE; OPTICAL METHODS OF CHEMICAL ANALYSIS. [R.F.C.]

Fluorine

Element number 9, fluorine, F, is the member of the halogen family of chemical elements that has the lowest atomic number and atomic weight. It is the most chemically energetic of the nonmetallic elements, and greatly exceeds both chlorine and oxygen in this respect. Because of this, it enters into combination with all the other chemical elements except the inert gases; and it is not found in nature in the elementary form. See HALOGEN ELEMENTS.



Uses. The uses of the compounds of fluorine date back to the time when man first learned to modify chemically the materials of his environment. Early metallurgy, ceramics, and glass technology were limited to materials that could be melted at the temperatures available. Since these temperatures are relatively low compared to those used in modern practices, it was a significant discovery to find a mineral that made melts and slags more fluid at the temperatures available. This mineral was called fluores and is now called fluorspar or fluorite. The name of the element is derived from the name of the mineral.

Fluorine-containing compounds are still used to increase the fluidity of melts and slags in the glass and ceramic industries. Fluorspar is introduced into the blast furnace to reduce the viscosity of the slag in the metallurgy of iron. Cryolite is used to form the electrolyte in the metallurgy of aluminum. Aluminum oxide is dissolved in this electrolyte, and the metal is reduced electrically from the melt.

The properties of the element are the basis for a variety of other uses of its compounds. The separation of the isotopes of uranium is accomplished by the gaseous diffusion of uranium hexafluoride. This process is difficult because the difference between the masses of the molecules containing the two major isotopes of uranium is only 3 parts in about 350. The fluorine compound of uranium is used in this process for two reasons: the hexafluoride is the only readily available volatile compound, and there are no naturally occurring isotopes of fluorine to complicate the separation process. The discovery of the catalytic use of hydrogen fluoride in organic chemical reactions in the laboratories of the Pennsylvania State University in the period 1933–1940 hastened the commercial production of this chemical. It was used to make aviation gasoline in World War II. The discovery of homologous series of fluorocarbons and their immediate derivatives at the Pennsylvania State University at about the same time created a class of new substances of considerable commercial interests. See ALKYLATION; POLYFLUOROOLEFIN RESIN; PROPELLANT.

Natural occurrence. Compounds whose molecules contain atoms of fluorine are widely distributed in nature. Many minerals contain small amounts of the element, and it is found in both sedimentary and igneous rocks. The chief ore is fluorspar (calcium fluoride), deposits of which are found in many parts of the surface of the earth. Large deposits of high-grade ore are not plentiful, but there are many low-grade or small deposits. Cryolite, Na_3AlF_6, which is used both in the ceramic industries and the metallurgy of aluminum, is found in quantities in only a few places. Fluorapatite, $Ca_{10}(PO_4)_6F_2$, frequently called rock

ered one of the rare ones. It is estimated that it comprises about 0.03% of the earth's crust. This makes it roughly as plentiful as nitrogen or chlorine, several times as plentiful as carbon, and many times as plentiful as copper or lead.

Properties. The element fluorine is a pale-yellow gas at ordinary temperatures. It liquefies at −187.99°C and freezes at −218.0°C. The odor of the element is somewhat in doubt. A very distinctive musty odor is detected whenever the element is present in the atmosphere. Since this odor is the same as that of oxygen difluoride, which results from a reaction of fluorine with water, there is some doubt as to whether the element itself has ever been smelled.

The chemical activity of the element is so great that it will react readily at ordinary temperatures with many other elementary substances such as sulfur, iodine, phosphorus, bromine, and most metals. Since the products of the reactions with the nonmetals are in the liquid or gaseous states, the reactions continue to the complete consumption of the fluorine, frequently with the evolution of considerable heat and light. Reactions with the metals usually form a protective metallic fluoride which blocks further reaction, unless the temperature is raised. Aluminum, nickel, magnesium and copper form such protective fluoride coatings. Fluorine reacts with considerable violence with most hydrogen-containing compounds such as water, ammonia, and all organic chemical substances whether liquids, solids, or gases. Hydrogen fluoride is one of the products of reaction with organic substances such as wood, paper, gasoline, and illuminating gas. The uncontrolled reaction of fluorine is accompanied by heat and light and produces *carbon tetrafluoride* as the chief carbon-containing product. Ingenious laboratory techniques have been employed to obtain from these reactions other products that contain both carbon and fluorine. Fluorine is so reactive that it will replace other nonmetallic elements from their compounds, even those nearest fluorine in chemical activity. It will replace chlorine from sodium chloride, and oxygen from silica, glass, and ceramic material. Fluorine can be stored only in containers made of metals that form protective fluoride coatings. It cannot be kept in contact with glass, wood, oil, or most ceramic materials. See ELECTRONEGATIVITY.

Production of the element. The preparation of the element can be accomplished satisfactorily only by means of an electric current. No chemical procedure has been found to produce the element in more than very small amounts. The essential feature of all electrical methods of producing fluorine is the electrolysis of liquid hydrogen fluoride that has been made conducting by addition of an alkali fluoride. Potassium fluoride is the salt used for this purpose and there are three temperature ranges in which potassium fluoride-hydrogen fluoride mixtures are liquid. The first isolation of the element by H. Moissan in 1886 employed liquid hydrogen fluoride held at a temperature below 0°C in a platinum vessel with platinum electrodes. In the range 230–320°C an electrolyte of the approximate composition of 1 mole of hydrogen fluoride per mole of potassium fluoride can be used to produce fluorine in iron or copper vessels using a graphite electrode for the anode. An American team of scientists discovered this during World War I. Great difficulty resulted from the failure of the electrical insulator at the high temperature in the presence of both elementary fluorine and hydrogen fluoride. Two hours was about as long as this operation could be continued before the insulation failed. In 1923 at the University of California, it was found that portland cement made an excellent insulator for use at this high temperature. This permitted the high-temperature fluorine-producing cells to be operated continuously around the clock. The temperature range preferred at present is approximately 100°C with an electrolyte of the approximate composition of 2 moles of hydrogen fluoride per mole of potassium fluoride. The anodes are either nickel or carbon, with the commercial units employing the latter.

The element has been used chiefly for laboratory experimentation. Many interesting compounds have resulted from such work. Mixtures of fluorocarbons were first made in the laboratory by the reaction between the elements fluorine and carbon. Other means are now employed to produce the fluorocarbons and fluorocarbon derivatives of commerce. The present chief use for the element is in the production of uranium hexafluoride. The second use is in the production of the halogen fluorides. These substances are used chiefly for experimental purposes in the laboratories. The compounds of fluorine are numerous, and they show a great variety of properties. Hydrogen fluoride is the most important of these compounds of fluorine just as water is of oxygen. It not only has important uses of its own, but also is the raw material for which many of the more important compounds of fluorine are prepared. See HYDROGEN FLUORIDE.

Inorganic compounds of fluorine. These can be divided into three kinds: the inorganic fluorides, the ionic compounds in which the fluorine exists in the negative ion in conjunction with some other element or elements, and the volatile nonsaltlike inorganic fluorides.

The extreme chemical properties of fluorine would be expected from its position in the upper right corner of the periodic classification of the elements. It has the possibility of being a constituent of more identifiable chemical species than any other element. This is because of its ability to combine with all other chemically active elements in great variety and also to form compounds with them of types not analogous to compounds free from fluorine. The difference between fluorine and the next most electronegative element, chlorine, is so large that the properties of fluorine-containing compounds cannot be predicted by analogy with the compounds of chlorine. A measure of the metallic or nonmetallic properties of an element is

indicated by its molal electrode potential. When this experimentally determined number is negative, it represents the nonmetallic properties. On this basis, fluorine (−2.85 volts) is farther away from chlorine (−1.36 volts) than chlorine is from hydrogen (0.00 volts). Examples of the great differences between fluorides and chlorides are the following: the chloride of silver is relatively insoluble in water but the fluoride is very soluble; the chlorides of barium, calcium, and aluminum are soluble in water, whereas the fluorides of the same elements are relatively insoluble; compounds of carbon and chlorine are relatively reactive and unstable but those of carbon and fluorine are relatively unreactive and stable.

Binary saltlike fluorides. Combinations of fluorine with all the metallic elements produce this type of compound. All the valence states of the metals are represented in the fluorides, and in addition, some in which the analogous chlorides are unknown, such as, vanadium pentafluoride and molybdenum hexafluoride. The fluorides are more prone to hydrolysis than are the corresponding chlorides, because of the great tendency of hydrogen fluoride (one of the products of hydrolysis) to combine with water. Silver fluoride, for example, cannot be obtained by removing the water from an aqueous solution of the salt. If this is tried, silver oxide and hydrogen fluoride are obtained. The binary saltlike fluorides show great tendency to combine with other binary fluorides to form a large number of complex or double salts. Cryolite, $3NaF \cdot AlF_3$, is one such compound.

There is a great variety of saltlike compounds in which the negative ion contains one or more other elements in addition to fluorine. The complex salts can be considered to belong to this classification. Cryolite, for example, can be considered as sodium fluoroaluminate. In this type of compound the element other than fluorine in the negative ion does not need to be metallic. Sodium fluosilicate and sodium fluophosphate are examples of such compounds. There are many phosphates in which the negative ion contains oxygen in addition to phosphorus and fluorine. The number of compounds of this kind already known is large but it is a very small fraction of the total number of such compounds theoretically possible.

Volatile inorganic compounds. These show the extreme properties of the element. The combination of fluorine with the nonmetallic elements always results in volatile compounds. The fluorine compounds with other halogens, IF_5, IF_7, BrF, BrF_3, BrF_5, ClF, and ClF_3, are all gases at room temperature or are readily vaporized liquids They are very reactive compounds. With each of the other nonmetals, such as oxygen, nitrogen, sulfur, and phosphorus, fluorine forms several volatile compounds, most of which are highly reactive. Boron trifluoride is a gas having unusual properties as a catalyst for organic chemical condensation reactions. The silicon fluorides are volatile. The best-known one, SiF_4, is the parent compound of the fluosilicates, when it is combined with metallic fluorides. The fluorides of all elements in which there are more than four atoms of fluorine per atom of the other element tend to be volatile. This applies even to the heavy metallic elements such as tungsten, osmium, and uranium.

Nearly all the volatile fluorides react with water to produce the oxyfluorides or oxides of the elements in addition to hydrogen fluoride. To this, however, there are several exceptions. If the central atom of the other element is of such a size that the fluorine atoms surrounding it have sufficient space to remain in the equilibrium spaces without crowding but, at the same time, do not have so much space that there are gaps between the fluorine atoms for another atom to squeeze through, the compound is resistant not only to hydrolysis but to many other chemical reactions. Sulfur hexafluoride is an example of such a compound. It is highly unreactive, although sulfur tetrafluoride hydrolyzes and undergoes other reactions readily. Carbon tetrafluoride is another example of a compound that is resistant to chemical reactions for the same reason. Its related compound, silicon tetrafluoride, reacts readily with water; but in this combination the central atom is larger, and therefore, the four fluorine atoms surrounding it do not protect it as well.

Organic compounds of fluorine. The fluorine-containing compounds of carbon can be divided into fluorine-containing hydrocarbons and hydrocarbon derivatives (organic fluorine compounds) and the fluorocarbons and their derivatives. The organic fluorides can be separated into the aromatic and the aliphatic classes The fluorine atom attached to the aromatic ring, as it is in fluorobenzene, is quite unreactive. In addition it imparts to the entire molecule great resistance to many reactions For certain purposes this is advantageous. Dyes, for example, that contain fluorine (or the CF_3 group which is equivalent to a fluorine atom) attached to the aromatic ring are more resistant

difficult to make and to keep and are not likely to become very important.

Fluorocarbons Compounds of carbon and fluorine that contain little or no hydrogen attached to carbon are called fluorocarbons. If they contain other elements, they are fluorocarbon derivatives. They are not organic compounds in that they do not react in the chemical reactions that have become the basis of organic chemistry. The reason is primarily that these reactions involve the hydrogen atoms attached to the carbon atoms. If there are a few such hydrogen atoms in a fluorocarbon molecule, the presence of a large excess of fluorine atoms deprives these hydrogen atoms of ease of reaction.

Because the carbon atom has the property combining with other atoms of carbon almost

out limit to form a large variety of structural patterns, there are many identifiable different molecular species of hydrocarbons and hydrocarbon derivatives The same property of the carbon atom provides for even more fluorocarbon and fluorocarbon derivatives, because these are considerably more stable than organic compounds so that more different structures can exist.

There is a still more numerous class of fluorine-containing compounds of carbon. Its molecules contain parts which are essentially fluorocarbon in nature and also parts which are organic or hydrocarbon in nature. There are potentially many millions of such hybrid molecules. The fluorocarbon parts retain their physical properties and chemical inertness, whereas the organic parts retain their chemical activity for organic reactions. Such compounds are useful for purposes demanding properties intermediate between the fluorocarbon and the organic. They are particularly useful in cases where the addition of an organic part to a fluorocarbon molecule can provide reactivity toward, or compatibility with, organic substances. This organic part provides means by which this molecule can be attached to some organic material such as the fibers of cloth or paper. This is advantageous since it provides the organic material with the desirable surface properties of the fluorocarbons.

The fluorocarbons and their derivatives are a relatively recent addition to the classes of chemical compounds. Despite earlier studies of the reaction of carbon and carbon-containing compounds, the simplest compound of the elements, CF_4, was obtained pure only in 1926; the second one, C_2F_6, was identified in 1930; and the third one, $CF_2{=}CF_2$, was prepared and identified in 1933. The demonstration of the fluorocarbons as an entire field of chemical substances by the isolation and identification of members of homologous series of substances came only in 1937.

In this work the fluorocarbons were prepared by the reaction of the elements fluorine and carbon with the help of mercury as a catalyst. Since this method produces a mixture of a very large number of gaseous, liquid, and solid substances which are difficult to separate, other means of producing individual compounds have been sought. Numerous complex chemical procedures have been devised. In one of these the element fluorine is caused to react with organic compounds with the aid of a catalyst and under exacting physical conditions. In another, the element fluorine is first used to produce metallic fluorides such as silver difluoride or cobalt trifluoride, and these are then caused to react with the organic substance. These two methods have produced many compounds in the laboratories but seem to be unsuited for large-scale industrial use.

One method of producing these substances that is currently used industrially starts with the production of C_2F_4. This is obtained from chloroform, CCl_3H, by a replacement reaction which produces CF_2ClH. This compound loses HCl upon heating, and $CF_2{=}CF_2$ is formed. This tetrafluoroethylene can then be caused to react in a number of ways. It combines with itself to form a useful polymer which is a chemically inert resin.

A more flexible and generally more useful industrial method of producing fluorocarbons and many derivatives is the electrochemical process. In this process an electric current is passed through a liquid mixture of hydrogen fluoride and some compound of carbon in a metal container in which there are metallic electrodes. The anode is nickel. Hydrogen is produced; and if the fluorocarbon product is gaseous, it escapes with the hydrogen and is separated therefrom. If it is liquid, it sinks to the bottom of the vessel and is readily removed, since it is insoluble in the hydrogen fluoride liquid. The fluorocarbon product frequently does not have the structure of the organic chemical raw material. The material to be used for a specified product is known only from experience. This process produces not only the fluorocarbons proper, whose molecules contain only carbon and fluorine atoms, but substances containing oxygen, nitrogen, and sulfur in addition to carbon and fluorine. There are also compounds containing small percentages of hydrogen in the products from the process.

The saturated fluorocarbons are resistant to chemical attack by other chemical reagents whether these be acidic or basic, oxidizing or reducing up to relatively high (about 600°C) tem-
[illegible] number of flu-

called unsaturated in organic chemistry, sometimes exhibit the properties of organic olefins and sometimes do not.

Fluorocarbon derivatives. The oxygen-containing fluorocarbons can be considered to be derived from OF_2 in which one or both fluorine atoms are replaced by carbon atoms. CF_3COF is a very reactive chemical but CF_3OCF_3 or higher molecular weight fluorocarbon oxides such as $C_3F_7OC_2F_5$ are almost as unreactive chemically as the parent fluorocarbons. Although of similar chemical formula, these compounds are chemically unlike the organic ethers. Compounds of the type $R_1R_2C{=}O$ are much more reactive than the fluorocarbon oxides. (R is used as an abbreviation of a fluorocarbon radical such as CF_3, C_2F_5, or C_3F_7.) Compounds of the type $R\overset{O}{\overset{\|}{C}}F$ produce the fluorocarbon carboxylic acids RCO_2H by reaction with water. These are nearly as strong acids as the common mineral acids. They can be reduced to aldehydes, $R\overset{O}{\overset{\|}{C}}H$, and alcohols, RCH_2OH. Thus, they become the starting point for the synthesis of the hybrid fluorocarbon-organic type of substance.

The nitrogen-containing fluorocarbons are derived from NF_3. Compounds of the class $R_1R_2R_3N$ are almost as resistant to chemical attack as the fluorocarbon oxides. They are not amines, since they do not react with acids to form salts. They are

properly called fluorocarbon nitrides. Compounds of the group R_1R_2NF and RNF_2 are also known. The next most important class of compounds is the sulfur-containing fluorocarbons. Compounds such as RSF_5 and $R_1R_2SF_4$ are known and are somewhat less resistant to chemical attack than the fluorocarbon nitrides. There are in addition many other types of fluorocarbon derivatives.

Properties of fluorocarbons. The fluorocarbons and their derivatives are chiefly of interest because of their physical properties. It is common to consider liquids at room temperature as being water-like or oil-like. The fluorocarbons provide a third class which is insoluble in the other two. This property of not dissolving oils or aqueous materials gives a fluorocarbon surface a valuable property. These other materials do not adhere to it. It resists being soiled by such substances. This is the basis of soil-resisting cloths and papers. The non-adhesion to other substances also provides a low coefficient of friction for other substances sliding on a fluorocarbon surface. Both liquid and solid fluorocarbons are potentially excellent lubricants.

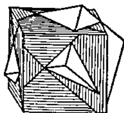
The fluorocarbons have very low surface tensions that are about one-half that of analogous hydrocarbons. A fluorocarbon molecule that has a polar group at one end for solution in water will act as a powerful surface-active agent, just as an organic fatty acid acts as a detergent or wetting agent. Addition of an organic carboxylic acid will reduce the surface tension of water by about one-half, and addition of a fluorocarbon carboxylic acid reduces it by half again.

The physical properties of the fluorocarbons and their inert oxygen, nitrogen, and sulfur derivatives combined with their resistance to chemical attack up to relatively high temperatures make them potentially superior substances for resins, plastics, oils, waxes, greases, elastomers, and fibers. *See* BROMINE; CHLORINE; FLUORIDE; FREON; HALOGENATED HYDROCARBON; HALOGENATION; IODINE [J.H.SM.]

Bibliography: J. H. Simons (ed.), *Fluorine Chemistry*, 2 vols., 1950–1954.

Fluorite

A mineral with composition CaF_2. It crystallizes in the isometric system. Crystals are usually cubes or penetration twins. The mineral is most commonly in coarse- to fine-grained aggregates. There is perfect octahedral cleavage. The hardness is 4, the specific gravity 3.18. Fluorite is colorless when pure but may be any color, or black due to impurities; it is commonly green, blue, yellow, or violet. Some but not all varieties fluoresce when exposed to ultraviolet light. Fluorite generally occurs in veins either by itself or as a gangue mineral associated with metallic ores. It also is found in cavities in limestone. The chief uses of fluorite are as a flux in the manufacture of steel and as the raw material in the preparation of hydrofluoric acid. [C.S.HU.]



Fluorite, showing twinned cubes. (From C. S. Hurlbut, Jr., *Dana's Manual of Mineralogy*, 16th ed., Wiley, 1952)

Fluorocarbon

An organic hydrocarbon in which all hydrogens have been replaced with fluorine. Fluorocarbons are extremely unreactive. Low-molecular-weight compounds dissolve in various organic solvents, but are in general quite insoluble. Fluorocarbon compounds are prepared from hydrocarbons by direct reaction with fluorine in the presence of a suitable moderator, by controlled fluorination with cobaltic fluoride, or by electrolysis of hydrogen fluoride solutions. Chlorinated hydrocarbons react with metallic fluorides (Swarts reaction) to form chlorofluorohydrocarbons, such as CCl_2F_2 (*see* FREON), which are quite inert and nontoxic. Although monofluorinated hydrocarbons are sometimes toxic and highly reactive, the completely fluorinated hydrocarbons have low boiling points, no toxicity, and great chemical stability. Extremely inert polymers are formed by polymerization of tetrafluoroethylene or chlorotrifluoroethylene. *See* HALOGENATED HYDROCARBON; POLYFLUOROOLEFIN RESIN. [A.L.H.]

Fluosilicate

The SiF_6^{2-} ion which is derived from fluosilicic acid, H_2SiF_6.

The fluosilicates are prepared from solutions of fluosilicic acid. The salts of barium, potassium, rubidium, cesium, and sodium are only sparingly soluble, whereas many of the heavy-metal salts are quite soluble.

Sodium fluosilicate, Na_2SiF_6, is used in the fluoridation of water supplies, in laundry scours, enamel frits, insecticides, and wood preservatives. *See* FLUORINE; SILICATE. [E.E.WR.]

Flutter, aeronautical

A problem in aeroelasticity, generally described as a self-excited oscillation, with a sustained or divergent amplitude, which occurs when a structure

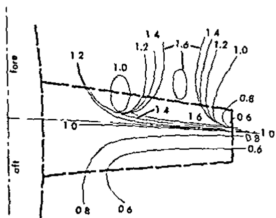
ture. If the flow remains laminar at all times, the problem lies in the domain of classical theory of aerodynamics and elasticity and is often called a classical flutter. If the flow becomes separated over part of the body or over part of the cycle, the resulting motion is called stall flutter. *See* AEROELASTICITY.

Flutter of an aircraft in flight usually causes catastrophic structural failure. Many lives and planes have been lost in the past because of flutter.

The basic cause of flutter is the extraction of energy from a flow by a moving object and the feeding of this energy into an oscillation of the object or a part of it. The mechanism of flutter is usually a mechanism of adjusting the relationship between the transient aerodynamic pressure and the transient wing displacement in such a way that, in certain specific modes of oscillation, the work done by the air balances or exceeds the energy that is dissipated by other sources of damping. It is important that the energy balance mentioned above be related to a flutter mode. Consider a bending-torsion flutter of a cantilever wing. When the flight speed slightly exceeds the first critical flutter speed, the energy exchange between the wing and the flow becomes such that the amplitude of the bending-torsion flutter oscillation increases with time, but in the meantime the wing drag is increased so that more power is required to maintain the flight.

In a harmonic motion, the phase shift between the force and displacement provides a mechanism for the energy exchange. Because the inertia forces and the elastic forces are conservative, the aerodynamic force is responsible for any possible flutter. The aerodynamic response to a wing motion depends on the dynamic pressure of the flow, the wing geometry, the amplitude of the motion, the mode of motion (which is governed by the geometry of the deflection surface), and the reduced frequency of the oscillation. It is through these dependences that the inertial and elastic characteristics of the structure exercise their influence, because the mode and frequency of motion is determined through the interaction of the inertial, elastic, and aerodynamic forces.

In modern airplane design, a flutter analysis is usually started at an early stage so that such decisions as the location of the engines and fuel tanks can be made. Such an analysis is necessary, as can be shown by an example. The diagram shows the results of a systematic study of the effect of engine locations on flutter speed by means of an electric analog computation for a typical fighter-type unswept tapered wing. A point on the contour curves of constant flutter speed represents the actual location of the center of gravity of the added mass, and the numbers shown refer to unity based on the flutter speed of the bare wing without any added mass. The diagram shows that the location of heavy masses on the wing is an important factor for design against flutter. A study of the efficient placement of masses with an objective of preventing flutter is called mass balancing. Mass balancing is often effective for the control of the control-surface flutter.



Contours of constant flutter speed for unswept tapered wing, showing effect of concentrated mass location using symmetric flutter.

An accurate determination of the flutter speed is of vital importance in airplane design, but it is also of great complexity and demands ingenuity and experience in analysis and experimentation.

To secure stability in the design speed range, the designer should (1) provide sufficient stiffness; (2) design aerodynamically so that flow remains unseparated in service condition (if aerodynamic forces are undesirable, as in a suspension bridge, attempts should be made to render the structure aerodynamically ineffective; drag reduction is especially beneficial in the case of stall flutter); (3) break the inertia and aerodynamic couplings either by a suitable arrangement of mass and elasticity distribution so that the elastic axis, the line of centers of mass, and the line of centers of aerodynamic pressure are as close to each other as possible, or by mass-balancing; and (4) provide servomechanisms to control the phase relationship between various components of motion. [Y.C.F.]

Flutter and wow

All sound recording and reproducing processes employ mechanical systems for the translation of electric waves or sound waves into and from the permanent record. If the speed of either or both the recording mechanism and reproducing mechanism vary at some rate, then a constant frequency impressed upon the recorder will not be reproduced as a constant frequency by the reproducer. The frequency of the tone will vary above and below the normal pitch depending upon the extent of the speed variation, and in frequency at the rate of the speed variation. The manifestation of this phenomenon is termed flutter. The term wow is used to designate this phenomenon for lower rates of flutter, such as those encountered in phonograph recording.

Flutter or wow is defined as the ratio of the root mean deviation in frequency of the signal to the average frequency expressed in per cent. Flutter or wow is given by

$$K = \frac{\Delta f_0}{\sqrt{2}\, f_0} \times 100$$

where K = flutter or wow, in per cent; f_0 = average frequency, in cycles per second; and Δf_0 = peak frequency deviation from the average frequency f_0, in cycles per second.

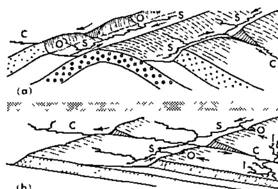
Flutter and wow are produced in disk recording systems by variations in the rotational speed of the recording and reproducing turntables and by a nonsymmetrically located center hole in the disk record. Flutter and wow are produced in tape and film reproducers by a nonuniform rotational speed of the capstan in the recording and reproducing machines, by nonuniform tension in the takeup and payoff reel systems, by the variety of idlers, and by a mechanical distortion of the tape in the recording or reproducing process. *See* DISK RECORDING; MAGNETIC RECORDING. [H.F.O.]

Fluvial erosion cycle

The sequential development of landscapes and their characteristic landforms by weathering processes and transportation of rock debris by running water. This cyclical development was proposed as the geographical cycle by William Morris Davis. An idealized landmass comprised of rock, exposed to processes of disintegration, decomposition, and transportation of constituent particles by currents acting under the influence of gravity, in time passes through sequential forms, eventually to attain near-flatness, a peneplain.

Stages in the fluvial cycle. The original surface becomes elevated at the start of a cycle. In the stage of incipient youth, small rills appear These deepen to become valleys or even deep canyons Local relief increases during youth, as valleys deepen, to a maximum at maturity, when most or all of the original surface is lost to erosion. The landmass then consists primarily of ridges and valleys with forms that depend on such factors as (1) kind of rock, whether massive as in the case of granite or stratified so that it erodes into cliffs and benches; (2) inclination of rock layers; (3) solubility of the rock; and (4) climatic conditions under which weathering and erosional processes occur. As the maximum possible relief depends on base level, or lowest limit to which streams can erode their valleys, the final stages of maturity mark the initiation of valley widening. Streams prevented from eroding vertically spend their energy impinging against valley walls. Debris moves down slopes to be transported away by river systems which develop during the cycle. Thus the divides are gradually lowered, so that as senility develops, toward the end of a cycle, landforms are low and subdued. Eventually the landmass becomes an almost featureless peneplain.

The drainage system at the initiation of a cycle consists of consequent streams which run down initial slopes. Prior to the appearance of structural control, insequent branchworks develop, with tributaries joining to form master (trunk) streams and rivers. To the extent that structure is expressed in the landscape as maturity is approached, subsequent streams develop valleys along lines of rock weakness, such as along fault zones or strata which are readily weathered and eroded. These may be quite straight, or curvilinear, depending on the rock structure present. Obsequent streams flow in



Genetic types of streams (a) In folded mountains. (b) On coastal plains. C, consequent; I, insequent, S, subsequent; O, obsequent. *(From A. K. Lobeck, Geomorphology, McGraw-Hill, 1939)*

a direction opposite to the original consequent streams of a region. They are usually short with steep gradients and often flow over escarpments. Most obsequent streams are tributary to subsequent streams.

Interruptions in the fluvial cycle. Interruptions and rejuvenations of a cycle may be caused by events such as climatic change or a sudden renewal of land uplift. A new cycle is initiated on the landscape of the earlier cycle, regardless of the stage of completion it had reached. Thus a mature region may experience dissection by processes which create a new set of youthful forms. Polycyclic landscapes may reflect more than a single rejuvenation of erosion cycles.

Though widely acclaimed and adopted, ideas advanced by Davis were vigorously criticized both in America and abroad. Walther Penck and other Europeans considered an initial uplift in a brief interval of time as unrealistic, proposing that a landmass could be elevated slowly or intermittently with the consequence that landscape evolution might differ materially from the suggested pattern. If intermittent, or if some of the weathering and transportation processes behave in ways not postulated by Davis, benches of relatively flat land, *Piedmonttreppen*, should flank mountain ranges. Within recent years, however, Europeans have generally adopted Davisian concepts and many hold to them with more tenacity than most American geomorphologists.

The idea that isostasy is a factor in structural geology was not incorporated in the theory of the

long continued It is thought that 10 ft or more of rock column must be eroded to reduce the elevation by 1 ft. The evolution of the youthful stage of the

sedimentary, rock section of mountains is thus lost at a rate much more rapid than that at which elevation is decreased, so that by the time an old-age stage is reached, rock exposed at the surface should consist of granite and metamorphosed materials characteristic of the basement complex. The length of time necessary to produce a peneplain across a broad region may represent most of the known geologic time scale. Shield areas, such as exist over much of Canada or Fennoscandia, may be the best approximations to peneplains in existence. See ISOSTASY.

For youthful topography the classic example of an initial stage is the bed of glacial Lake Agassiz, North Dakota, where little stream cutting has been experienced by the lakebeds. A more advanced stage occurs in the Colorado Plateaus, which have developed deep canyons, foremost among which is the Grand Canyon of the Colorado, but where widespread areas of original surface remain. Much of the Allegheny Plateau has reached maturity. Small remnants of original surface remain, but most of the area consists of valley slopes or bottoms. The Olympic Mountains in Washington and many other mountains are maturely dissected.

The supposition that lateral corrosion accounts for valley widening during the mature stage of the

uct of valley filling, rather than of erosional attack on valley walls. This is shown by borings through the surface alluvium which ordinarily tops a deep section of valley fill, particularly along the larger valleys. The deposits are not the typical thin veneers which overlie horizontal bedrock as planed by the erosional work of laterally migrating streams. Nor are extremely wide flood plains, such as those of the Lower Mississippi River or San Joaquin Valley of California, evidence of old age. Though Davis realized that landforms are both erosional (destructional) and constructional in origin, he seldom accorded more than passing mention to alluvial forms in his consideration of the normal cycle of erosion. See FLOOD PLAINS.

Other geomorphic cycles. The fluvial erosion cycle sets patterns for the recognition of cycles (1) in arid climates, where the presence of *interior* basins introduces a complicating factor; (2) under glacial conditions, where ice is the dominant agent of transportation; (3) on coasts, where criteria have been postulated for recognition of young, mature, and old stages; (4) for limestone *regions*, where rock solubility accounts for karst features; and (5) for other special conditions, such as *in* tropical, mediterranean, and other climatic regions. See COASTAL LANDFORMS; DESERT EROSION FEATURES; FLUVIAL EROSION LANDFORMS; GLACIATED TERRANE; KARST TOPOGRAPHY. [R.J.RU.]

Bibliography: W. M. Davis, The geographical cycle, *Geographical Journal*, XIV(5):481–504, 1899; W. M. Davis, *Geographical Essays*, reprint, 1954.

Fluvial erosion landforms

Landforms that result from erosion by water running on the earth's surface. This water may concentrate in channels as streams and rivers or run in thin sheets down slopes. All land surfaces, even those in the driest deserts, are subject to modification by running water.

Factors affecting fluvial erosion. Running water erodes in the following ways.

1. Direct lifting. Most stream flow is turbulent. The whirling eddies of turbulent water can dislodge material and carry it downstream.

2. Abrasion and impact. The solid particles carried by the stream can wear down bedrock portions of the stream channel or the large rock fragments that pave the bottom. The impact of rock particles thrown against rock may also knock loose fragments which in turn are carried off by the stream.

3. Cavitation. At stream velocities of 25–30 ft/sec cavitation can be effective. In this process vapor bubbles in the water suddenly collapse. This collapse theoretically generates sudden pressures up to 1500 or 2000 lb/in.^{2} Thus bubbles which collapse against a portion of the stream channel act as a sort of hammer blow and provide an effective erosive force.

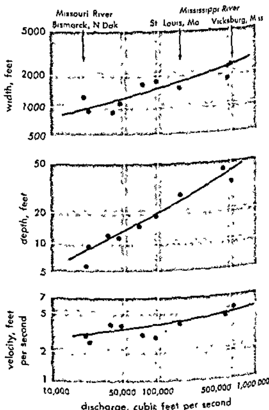


Fig. 1. Width, depth, and velocity of streams increase downstream with increasing discharge, as illustrated by the Mississippi-Missouri river system. (From L. B. Leopold and T. Maddock, Jr., *Hydraulic Geometry of Stream Channels and Some Physiographic Implications, USGS Profess. Paper 252*, 1953)

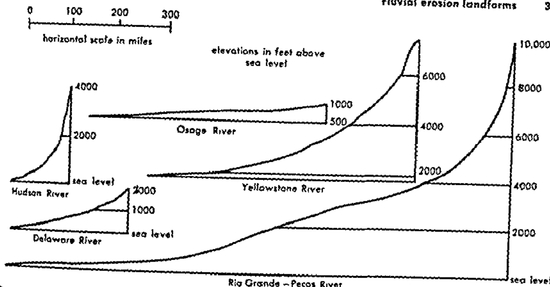


Fig. 2. The profile of a river decreases in gradient from head to mouth and is concave upward. (From H. Gannett, Profiles of Rivers in the United States, USGS Water Supply Paper 44, 1901)

4. Solution. Enormous amounts of rock material are carried annually to the oceans in dissolved form. Some solution may take place in the stream channel, but most dissolved material probably is contributed to streams by underground water.

5. Undercutting. A stream that erodes laterally by one or more of the above processes may undercut its channel bank or valley wall to the extent that gravity causes material to slump into the river where it is moved downstream.

Once in a stream, material may be carried along in solution, buoyed up in suspension by turbulent water, or rolled, slid, or bounced along the channel bottom.

Stream profile. Streams erode, transport, and deposit material in an orderly fashion. Each stream and its tributaries forms a system that is in dynamic equilibrium. For instance, the width and depth of the channel, as well as the velocity, change in a predictable fashion with changes in the amount of discharge. This is true for both changes of discharge at a particular spot on a river and increases in discharge downstream.

Figure 1 shows that if a stream's depth, width, and velocity are measured at a standard flow frequency (here median flow) at several points along the stream, then, as discharge increases downstream, the width and depth increase as expected. But surprisingly velocity also increases downstream. This happens to satisfy the relation that discharge must equal width times depth times velocity. Erosion increases the stream's depth and width but not rapidly enough to accommodate the discharge at a constant, much less a decreasing, velocity.

A corollary to the increase in width, depth, and velocity of a stream as discharge increases downstream is the over-all shape of the longitudinal profile of a stream from its headwaters to the mouth. This profile is parabolic in shape and concave to the sky, and its gradient decreases rather than increases downstream (Fig. 2).

Base level This general concept is useful in considering the erosional activity of a stream. A base level is the point below which a stream cannot erode. Thus the point at which a stream enters the ocean is the ultimate base level of the stream. If the ocean level is constant, the stream, over the course of years, will reduce its profile toward this base level, although it will reach it only at its junction with the ocean. As a result, although the profile will become gentler with time, it will always maintain its parabolic form, concave upward. If base level is lowered, either by falling sea level or rising land, then erosion will take place as the stream seeks to adjust itself to the new, lower base level. Conversely, a rise of base level, by whatever means, tends to cause stream deposition. Although the ocean is the ultimate base level, other local base levels may exist along a stream's course. Thus the presence of resistant rock will provide a local, although temporary, base level for all segments of the stream upstream from it. Other temporary base levels include lakes and swamps along the stream course For a tributary stream, base level is the elevation of the trunk stream which it enters (Fig. 3).

Differential erosion. This aspect of erosion is critical in considering stream erosion. Earth materials erode at different rates. Thus quartzite is more resistant to erosion than shale. Climate may also affect the rate at which rocks are eroded. Limestone, for instance, is soluble in natural waters and therefore can be removed more rapidly in well watered climates than in arid or semiarid regions. As a result limestones often underlie valleys in eastern United States, whereas they form ridges in the drier, western states.

Features of fluvial erosion. Features of river erosion range in size from minute features of the

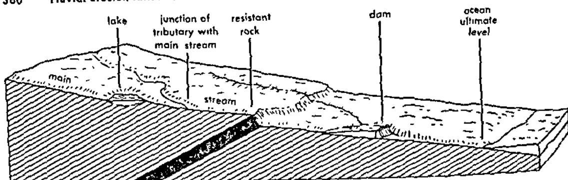


Fig. 3. Differing base levels for streams. (From L. D. Leet and S. Judson, *Physical Geology*, 2d ed., Prentice-Hall, 1958)

channel to regional landscapes. The features that are most intimately related to fluvial erosion are undoubtedly those that are directly associated with the stream channel. Among these are included not only the cross section of a stream channel but also the long, concave-upward profile of a stream from head to mouth. Even when this shape is not directly related to a stream channel, it may still be due to running water which flows down slopes as sheets or in broad shallow rills during exceptional rains.

Stream terraces. A stream terrace is a relatively flat surface extending along a valley with a steep bank separating it from a lower terrace or from the flood plain of a river. The flat surface of the terrace represents a remnant of the stream's former flood plain. The terrace now stands above the general level of the stream's new flood plain because the stream has eroded downward and abandoned its earlier flood plain (Fig 4).

Waterfalls and rapids. These are impressive, but geologically transitory features of a stream. They are present because of a sudden drop in the stream course. This interruption in the profile of the stream is in many instances due to resistant rock

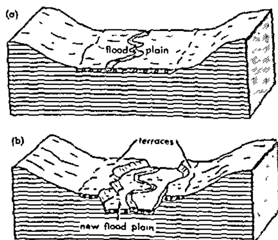


Fig. 4. Development of stream terrace. (a) Original flood plain. (b) Terrace marks level of abandoned flood plain above level of new flood plain.

which the stream cannot easily remove. Thus Niagara Falls is held up by a thick layer of resistant dolomite which overlies less resistant shale. Other falls may exist because a main valley has been deepened much more rapidly than a side valley, leaving tributary streams to plunge into the main valley over a falls. Several of the falls in the Yosemite National Park in California are of this type. The main valley of the Merced River has been deepened by now-vanished glacier ice to a level much below less severely eroded side valleys. Whatever the particular cause of waterfalls, the water which moves over them works constantly toward their destruction. As the falls erode, they tend to migrate upstream as suggested in Fig 5. Rapids may form as falls are destroyed or because of a rapid increase in the stream gradient.

Stream patterns. These are varied and each pattern indicates something about the underlying earth materials. A stream system may describe a branching, treelike or dendritic pattern. Such an arrangement of a stream and its tributaries is actually a random orientation of channel ways and develops on earth materials that have a relatively uniform resistance to erosion, such as granite or flat-lying sedimentary rocks.

When the subsurface rocks are not uniform in resistance. streams tend to cut their valleys in the zones of least resistance. Therefore the resulting stream patterns tend to follow the underlying rock pattern. As an example, bedrock fractured into rectangular blocks tends to produce a rectangular drainage pattern reflecting the easily eroded fracture zones. A trellis pattern also develops because of varying resistance to erosion of the underlying rock. In this case parallel bands of resistant and nonresistant rock are etched out by streams carving valleys in the least resistant bands. These streams often join a master stream which cuts across the resistant and nonresistant beds alike, giving the entire drainage the aspect of a garden trellis. Sometimes the stream pattern reflects the original topography of an area. Thus, on a domal area, streams tend to radiate outward from the high center; the term radial applies to this drainage. See RIVER.

Summary. Running water is the most important of all the processes which fashion the landscape.

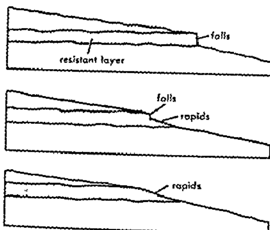


Fig. 5. Falls formed where stream crosses resistant rock. As falls retreat upstream they may develop into rapids.

Directly or indirectly, water flowing in streams is responsible for carving most of the valleys of the continents. Streams cut into the land, eroding both downward and from side to side. These incisions allow gravity to operate along the valley walls. This motion of material under the influence of gravity may be either rapid, as in landslides, or slow, by almost imperceptible creep of the unconsolidated surficial cover derived by weathering from the firmer rock beneath. In addition to direct stream erosion and indirect contributions by gravity, material comes to the stream by sheets of water flowing down slopes. No matter how streams acquire their load of material, they serve as endless conveyor belts moving debris from the land to the oceans. As an area is worn lower and lower, its landscape passes through a series of stages known as the fluvial erosion cycle. See FLUVIAL EROSION CYCLE.

[S.J.]

Bibliography: W. D. Thornbury, *Principles of Geomorphology*, 1954.

Fluxmeter

An instrument designed to measure magnetic flux (see MAGNETIC FLUX). The fluxmeter is essentially a moving-coil ballistic galvanometer with a long period and restoring torque reduced to minimum; it is heavily overdamped in a low-resistance circuit (see GALVANOMETER). These conditions are necessary in order for the instrument to measure the electromotive force (emf) impulse $\int e\,dt$ without appreciable error, regardless of the length of time of the impulse. The low restoring torque (ideally zero) may be obtained by substituting a silk fiber for the usual metallic strip and making connection to the moving coil by light metallic spirals designed to exert negligible control of the motion.

A search coil is connected to the fluxmeter, as shown in the illustration, and the change in the flux that produces the induced emf occurs within the search coil (see SEARCH COIL). This change may be caused by a change in the flux that links a stationary coil or by a motion of the search coil in the field so that there is a change in the flux linkage of the coil.

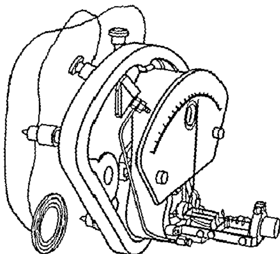
In operation of the fluxmeter, the change of flux causes the coil to move to a deflected position. The angle θ through which the coil turns is proportional to $\int e\,dt$.

$$G\theta = \int e\,dt$$

where the proportionality factor G is known as the motor constant of the galvanometer. But $\int e\,dt = N\,\Delta\Phi$, where $N\,\Delta\Phi$ is the total change in flux linkages in the search coil having N turns. Hence,

$$G\theta = N\,\Delta\Phi \qquad \Delta\Phi = \frac{G\theta}{N}$$

The ballistic galvanometer must be properly calibrated by determination of the value of G for the conditions under which the instrument is to be used. The flux-density sensitivity of the fluxmeter



Fluxmeter. (From F. A. Laws, Electrical Measurements, 2d ed., McGraw-Hill, 1938)

depends upon the value of the flux density within the air gap of the galvanometer. The sensitivity will be greater for weaker fields in the gap.

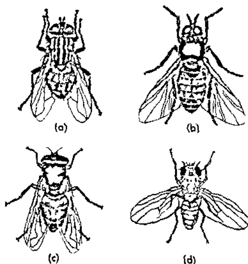
[K.A.M.]

Bibliography: F. K. Harris, *Electrical Measurements*, 1952; F. A. Laws, *Electrical Measurements*, 1938; W. C. Michels, *Electrical Measurements and Their Applications*, 1957.

Fly

Any member of the insect order Diptera. Flies are characterized by a single pair of membranous, netveined wings, laid flat on the back when at rest; by a second pair of wings that are rudimentary structures, stalked and tipped with small knobs called halteres; by mouthparts for piercing and sucking, lapping, or sponging, but never biting; by complete metamorphosis; and usually by a larva of the maggot type.

There are about 85,000 species in this world-wide and highly successful order. Besides the many in- sects commonly called flies, there are several o



Common flies. *(a)* The housefly, *Musca domestica;* length about ⅜ in. *(b)* The black fly, *Simulium* sp. *(c)* The rat-tailed maggot, *Eristalis tenax*, length of drone fly ¾–1 in. *(d)* The fruit fly, *Drosophila melanogaster;* length about ³⁄₁₆ in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

that are true flies in the entomological sense, such as the mosquitoes, gnats, and midges. However, various insects in which the term fly is employed as a suffix (dragonfly and stonefly) are *not true flies.*

Importance. Included in the flies are some of man's most annoying and serious enemies. In addition to mosquitoes, other biting flies include the gnats, black flies, deer flies [illegible] riers of [illegible] tsetse fl[illegible] an important vector of the organism that causes tularemia. Well-known enemies of domestic animals include the horseflies, bot flies, warble flies, and louse flies. As agricultural pests the Mediterranean and Oriental fruit flies, Hessian fly, clover seed midge, and apple maggot rank high on the list.

The robber flies are predaceous insects; many of them prey upon pest insects and are thus beneficial, whereas others prey actively upon honey bees and thus are harmful species. Some of the robber flies so closely resemble bees or bumblebees that they are recognized as flies only by trained observers. The bee flies also appear more like bees than like flys and are commonly misidentified.

A number of flies are parasitic on other insects. One family, the leaf miners, develop inside leaves, the tiny larvae leaving characteristic irregular, curving, pale paths to mark where they have eaten. The miners seldom do serious damage to mature plants, but sometimes may materially injure tender plants such as tomatoes and peppers.

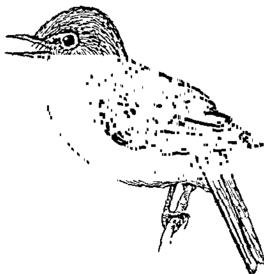
Flies as larvae and adults are a major source of food for many wild animals that depend upon insects. The birds, bats, rodents, reptiles, and amphibians that prey upon insects are major factors in keeping the highly prolific flies in check.

A number of flies, such as the flower flies and bee flies, are valuable as pollinators of plants, and are second only to bees in this important work.

Life history. The life histories of flies are as diverse as their appearance and habits. The egg usually hatches into a legless, wormlike, usually white larva called a maggot. There is a resting pupa, followed by the adult. In some species, such as the fruit flies of the genus *Drosophila*, which are common around decaying fruit and widely used in genetics research, the complete cycle may be accomplished in 8 days; in others there may be but a single generation a year. In the midges the aquatic larvae are red, and are called blood worms rather than maggots. Mosquito larvae likewise differ from the typical maggot pattern, as do the larvae of some other flies. *See* DIPTERA. [J.D.B.]

Flycatcher

Any member of the perching-bird family Tyrannidae, a highly successful and diversified American family with 118 genera and 365 species. Thirty species are found in the United States. Most flycatchers are shy, gray to olive birds known only to those who actively search for them. However, a few familiar to almost everyone include the gaudy scissor-tailed flycatcher, the kingbird, and the common phoebe. They all take their food on the wing, launching out from an elevated, exposed perch at the sight of an insect, much as a hawk hunts larger prey. They are solitary, spending most of their time alone on hunting stands. The United States species are all strongly migratory; of them, the eastern phoebe is the only one to winter north of southern Florida. Flycatchers do not sing as much as most other perching birds, for they have a rather poorly developed syrinx. Nevertheless, the calls of many species are melodious and pleasing. *See* KINGBIRD; PASSERIFORMES; PEWEE; PHOEBE. [J.D.B.]



The Acadian flycatcher, *Empidonax virescens;* length to 6¼ in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Flying fish

Any of about 65 species of marine fishes, mostly living in tropical waters, belonging to the family Exocoetidae. Flying fishes are so named because of their habit of leaping from the water when pursued by predators and sailing through the air for some distance, or sometimes simply skipping along on the surface. Some species are capable of flying 200 yards or more, but rarely rise over 4 ft above the surface of the water. They are fishes of the open sea, and travel in large schools. Both herbivorous and carnivorous species occur, ranging in length from 2–18 in. They have large, winglike pec-



The flying fish, *Parexocoetus mesogaster;* length 7 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

toral fins set high on their sides, large pelvic fins, and a rather large, deeply forked caudal fin with the lower lobe larger than the upper. The tail provides the driving force which thrusts the fish from the water, and the glide is maintained by spreading the paired fins. *See* ACTINOPTERYGII. [J.D.B.]

Flying saucers

A household phrase that came into use shortly after June 24, 1947, when Kenneth Arnold, a Seattle businessman, flying his plane, glimpsed a series of bright objects in the vicinity of Mt. Rainier, Washington. The reflected sunlight from their surfaces suggested that they might have been metallic. The enormous speeds of the objects and their peculiar wobbling motion that gave them the appearance of saucers skipping over a pond particularly attracted Arnold's attention. Each appeared larger than an airplane and crescent in form.

Numerous reports. The story received wide publicity. Soon observers scanning the sky reported all kinds of extraordinary objects coursing across the heavens, luminous objects by night and nonluminous ones by day. Some seemed to hover for hours; others appeared to flash by at incredible speeds. Observers inferred rapid motions from the sudden disappearance of the objects. Many shapes but chiefly disks or saucers were reported, hence the term flying saucers.

Night observations produced even more spectacular phenomena. There were glowing fireballs, some moving slowly, others flashing at incredible speeds. Green flares, reported from the southwestern United States, showered sparks and left persistent luminous trails. Airplane pilots reported the appearance of glowing disks that buzzed their ships and trailed them, apparently just beyond the wing tips.

Earlier records. These observations were not unprecedented. Early in the twentieth century, Charles Fort made a hobby of collecting accounts of strange events from varied sources, chiefly newspapers. He published these unconfirmed reports as books, thus giving permanent circulation to a hodge-podge of unexplained reports and to his speculation that a nearby outer world was peopled with intelligent beings who visited Earth. The Fortean Society was formed to perpetuate his ideas.

Earlier a great furor arose in 1897 over observations of a seemingly cigar-shaped object that, from reports, was assumed to be moving eastward from San Francisco. Newspapers reported the craft near Chicago. Many testified to having seen the ship and its searchlights; thousands of watchers lined the lake front and scanned the sky. Few heeded when astronomers advised that the searchlight was only the bright star Betelgeuse. Innumerable hoaxes further confused the issue. Thus, in many respects, the mythical airship of 1897 paralleled the flying saucers of today.

Public confidence. Certain differences do exist, however; in 1897 T. A. Edison asserted that an airship did not exist. So great was the respect for his authority that, almost overnight, reports of the ship vanished from the newspapers.

Recent reports have been received differently. The U.S. Air Force, with responsibility for investigation of reported sightings, classified the subject as secret. The public is fully aware that many scientific activities are under secret classification and has a still-vivid recollection of wartime adversities glossed over in press releases. Against these experiences, the public is uncertain how to interpret the repeated statement from the Air Force that flying saucers present no threat to national security and that there is no evidence of their extraterrestrial origin.

The designation unidentified flying objects (UFO) is a misnomer because it implies that the reported sightings are of material reality.

To enhance the confusion, some writers have charged the Pentagon with suppressing information that proves the spatial origin of the objects. These critics declared that the government feared the panic that release of such information might cause and was, therefore, withholding facts.

Utter nonsense!

The Air Force carefully screened the reports and found most of them too fragmentary for analysis; others can be explained in terms of known natural phenomena.

Phenomena responsible for effects. Even the most reliable and sensational saucer reports can be explained reasonably and logically in terms of one of the following natural phenomena.

Light reflections from material objects account for most reports of saucers. These include reflections from planes, wind-borne paper, kites, weather balloons, sky-hook balloons (the large ones that attain altitudes in excess of 100,000 ft), feathers, milk-weed seeds, and spider webs. The planet Venus has been responsible for many reports.

times moving cirrus clouds produce the illusion of rapid motion.

At night, any of the brighter stars or planets may be responsible. This is particularly true when the object lies almost on the horizon and is distorted by intervening layers of air. Several years ago a bright saucer that buzzed an Air Force plane in which I was flying proved to be a mirage of Sirius, the brightest star in the sky. Mirages are a common cause of the sighting of saucers.

Many luminous objects, such as landing lights of airplanes, lights carried by weather balloons, fireflies, reflections of searchlights or other light sources on cloud layers, or flying birds, can produce weird effects. Owls or other nocturnal birds that live in hollow trees occasionally appear self-luminous because they are powdered with foxfire, the fungus that makes rotten wood glow in the dark. Meteors and other fireballs flashing across the sky can be spectacular. Some have a remarkable green hue, caused by magnesium that glows when the piece of rock is heated by friction with the Earth's atmosphere.

Sundogs or mock suns and their associated halos have been the source of many reports, especially from airline pilots. Bright patches, caused by reflections or refractions of sunlight or moonlight from frost crystals, can follow the observer much as a rainbow appears to move. Fog, haze, and vapor trails are additional sources.

Much has been made of saucers detected by radar or in simultaneous radar and visual sightings. Radar beams can be bent earthward or scattered by atmospheric irregularities just as light waves can. These effects were known and understood long before the radar sightings.

Some people claim direct contact with flying saucers either by riding in one or by talking with an operator of one. None of the many books and articles of this type carries the slightest aura of authenticity. None deserves serious consideration.

There is one lesson to be drawn from the pattern of reported saucer sightings. Transient phenomena are always present, ready for the imagination to build on. Any publicity that increases the number of persons looking at the sky will also augment the number of reported sightings. See HALLUCINATION; METEOR; SAINT ELMO'S FIRE; SUNDOG.

[D.H.M.]

Bibliography: D. H. Menzel, *Flying Saucers*, 1953.

Flywheel

A rotating mass used to maintain the speed of a machine between given limits while the machine releases or receives energy at a varying rate. A flywheel is an energy storage device. It stores energy as its speed increases and gives up energy as the speed decreases. The specifications of the machine usually determine the allowable range of speed and the required energy interchange.

Theory. The energy, speed, and size of a flywheel are related by the equation

$$Wk^2 = \frac{182.4gE}{n_1^2 - n_2^2} \quad (1)$$

in which W is weight in lb, k is radius of gyration in ft, g is the gravity constant in ft/sec², E is the energy change in ft-lb, and n_1 and n_2 are the maximum and minimum speeds in rpm. The term $n_1^2 - n_2^2$ is sometimes replaced by the product $2c_f n^2$ to give

$$Wk^2 = \frac{91.2gE}{c_f n^2} \quad (2)$$

in which

$$c_f = \frac{n_1 - n_2}{n} \quad (3)$$

is the coefficient of fluctuation and

$$n = \frac{n_1 + n_2}{2} \quad (4)$$

is the average speed.

The energy interchange cycle of a flywheel may occupy any amount of angular motion about its center of rotation. Most frequently it is a cycle of one or more revolutions, but in internal combustion engines with multiple cylinders, it may be only a small part of a revolution.

All rotating components in a machine contribute to the flywheel effect and should be considered in a complete analysis of the design. The term Wk^2 in Eqs. (1) and (2) is the flywheel effect and may be made up of the sum of the many rotating weights and their respective radii of gyration squared. Equation (2) can therefore be written in more general terms.

$$W_1k_1^2 + W_2k_2^2 + \cdots = \Sigma Wk^2$$
$$\Sigma Wk^2 = \frac{91.2gE}{c_f n^2} \quad (5)$$

Flywheel energy. In the design of a flywheel the first step is to determine the energy interchanged. This can be done best by plotting the torque required by the machine against the angular motion of the flywheel for one cycle of θ_0 radians (Fig. 1). In this plot, the area under the torque curve represents the total energy requirements of the machine during the cycle. Then, average torque T_{av}

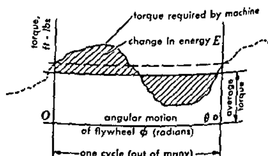


Fig. 1. Graphical determination of energy change for a flywheel.

Coefficients of fluctuation

Machine	C_f
Crushers, hammers, punches, shears	0.100–0.200
Machine tools	0.020–0.030
Electric generators	0.003–0.005
Compressors and pumps	0.040–0.050
Textile machinery	0.010–0.030

defined as the total energy of the cycle E_0 divided by the length of the cycle θ_0 in radians is drawn on the same torque scale. The areas between the original torque curve and the average torque curve above and below this average torque (shaded areas on the drawing) should be equal and are the E, or change in energy, in Eq. (1).

The second step is to establish the speed range, n_1 and n_2, of the machine or to select the appropriate coefficient of fluctuation from the table.

The third step is to calculate the Wk^2 required by substituting the values of E, c_f, and n previously determined. From this total Wk^2 required is subtracted the values of Wk^2 for the several rotating components in the machine if these are significantly large. The remaining Wk^2 must be supplied by an appropriate flywheel or flywheels.

The designer can pick any number of values of W and k whose quotient Wk^2 will be the required amount. The limiting factors are weight and diameter. A suitable decision is based on space availability and weight restrictions. It is sometimes necessary to gear the flywheel to a higher speed than the machine to reduce Wk^2 so that weight and size restrictions can be met.

Design of flywheel. The difficulty of casting stress-free spoked flywheels, leads the modern designer to use solid web castings or welded structural steel assemblies. For large slow-turning flywheels on heavy duty diesel engines or large mechanical presses, cast spoked flywheels of two-piece design are standard practice (Fig. 2)

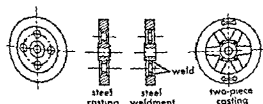


Fig. 2. Typical flywheel structures.

Because it is difficult to calculate the stresses in a spoked flywheel, these are usually designed on the basis of a maximum allowable peripheral speed of 5000–7000 feet per minute for cast iron. The higher value is for rims cast by special techniques to eliminate rim porosity or blow holes. Higher rim speeds to 10,000 fpm may be used with cast steel. Structural steel plate welded flywheels of the solid web type can be accurately analyzed for stress by applying the equations of rotating disks to the components of the flywheel and allowing a factor for the stresses caused by welding. [L.S.L.]

Bibliography: C. Carmichael (ed.), *Design and Production, Kent's Mechanical Engineers Handbook*, vol. 2, 12th ed., 1950; T. B. Jefferson and W. J. Brooking, *Mechanical Design*, 1951; V. L. Maleev and J. B. Hartman, *Machine Design*, 3d ed., 1954; A Valance and V. L. Doughbie, *Design of Machine Members*, 1951.

Foam

A material made up of gas bubbles separated from one another by films of liquid. The bubbles are spherical when the liquid films separating them are thick (approximately 1 mm); they have a polyhedral shape when the liquid films between them are thin (approximately 0.01 mm). Pure liquids do not foam, that is to say, they cannot produce liquid films of any permanence. Relatively permanent films are created only when a substance is present that is adsorbed at the surface of the liquid. Substances capable of being so adsorbed may be in true solution in the liquid or may be particles of a finely divided solid, which, because of poor wetting by the liquid remain at the surface. In both cases, surface layers of the added substance are produced. The reluctance of the adsorbed substance to enter the bulk of the liquid preserves the surface and hence, the thermodynamic stability of the foam. See SURFACE ACTIVE AGENT.

Although thermodynamically stable, a foam is mechanically fragile. Offsetting this fragility to some extent are two mechanisms that provide the liquid films with resiliency and plasticity.

Resilience of films. The resilience, or elastic recovery after external mechanical strain, of a film of liquid originates in the higher surface tension of freshly formed liquid surfaces as compared to that of older surfaces. The higher surface tension of new surface heals the incipient breach by drawing in toward it the thicker portions of the film. This property depends on the presence of sufficient bulk, or subsurface, solution. The effects of impaired resilience, caused by extreme depletion of underlying liquid from old films by prolonged drainage, is evident in their brittleness. A lack of resilience causes the ultimate collapse of foams.

Willard Gibbs deduced the existence of surface elasticity E from a priori considerations and gave it quantitative definition as

$$E = 2s\,\frac{d\gamma}{ds}$$

where s is the surface area of the film, and $d\gamma/ds$ is the surface-tension gradient. Gibbs based his deduction on the thermodynamic theorem that a solute that lowers surface tension has an excess concentration at the surface. Any stretching of this surface, which would form new surface from material in the bulk liquid, is attended by a deficiency of surface-active material at the area affected and, consequently, results in a higher surface tension than that of the surrounding surface. A surface-

tension gradient, $d\gamma/ds$, is produced in this way.

In thick films, however, the surface elasticity E, as described in the equation above, is transient. The surface-tension gradient due to the new surface can be reduced spontaneously in two different ways: (1) the surface layer (or excess surface concentration) diffuses from areas of lower to those of higher surface tension (or from higher to lower surface concentration); (2) the solute in the interior of the film adsorbs at the surface. Both phenomena occur simultaneously, although at different rates. Only the first phenomenon is effective, however, in restoring a stretched film, at least partially, to its original thickness. This occurs by the following mechanism: the flow of the surface layer pulls along with it an appreciable amount of underlying liquid, which restores the thickness of the thinned area. The effective elasticity of a liquid film, therefore, is given by a comparison of the rates at which these two phenomena take place, rather than by a calculation of E from the equation given by Gibbs. If the first phenomenon occurs more rapidly than the second, the film restores itself automatically after being stretched; if the second is more rapid, the film will lack resilience. Measurements of the rates of flow of monomolecular films of oleic acid on water show that the monolayer carried with it a quantity of subsurface water approximately 30 μ (microns) deep and that it moved with a speed of 2-5 cm/sec. The surface layer would, therefore, move to cover a distance of 1 mm in 20–50 milliseconds. During the same period of time, the surface tension of a newly produced surface of, for example, 0.4% sodium oleate solution would not yet have reached static surface tension equilibrium. Were the extension of the surface much greater, however, diffusion to the surface from the interior would have reached equilibrium before the lateral movement of the surface layer had had time to cover the new surface. This calculation, at least, confirms the general experience that soap films cannot be stretched indefinitely without losing resilience.

The above-mentioned mechanism for resilience of liquid films is not enough by itself to impart exceptional time-stability to foam. The gravitational flow of the liquid out of the films soon renders them so thin and so concentrated that they cannot create an area of higher surf[illegible]

[illegible]

[illegible]

of the films is retarded, however, where surface plasticity is a property of the solution; such solutions, although perfectly fluid in their flow properties as bulk liquids, have plastic surfaces; that is, the surface acts as if it were a thin sheet of a soft plastic material that can resist small shearing stresses and retain its original form intact until the applied shear is great enough to cause plastic flow. By virtue of this property, the film can resist any stretching or displacement of the surface in response to local stress. In addition, since the plasticity has a depth of 10–40 μ, a soap film, which is plastic at both surfaces, may contain very little water that is free to drain out; the films then remain thick and the foam is correspondingly long-lived. Everyday examples of films possessing surface plasticity are whipped cream and whipped egg-white (meringue). In these, the foams are stabilized by surface-plastic foam stabilizers. In ready-whip shaving cream, a mixture of soaps or detergents acts as a foam stabilizer. The time-stability of such foams is enhanced enormously compared with those that lack surface plasticity.

Elimination of foam. Although foams of exceptional stability are desired in some commercial applications, foam is a nuisance in many situations. In the glue and paper industries, in the processing of foodstuffs, in sewage disposal, or in the dehydration of oil, foams cause undesirable complications. A common recourse is the addition of chemical antifoams, which are usually insoluble liquids of very low surface tension. When a droplet of such a liquid is sprayed onto the foam or is carried into it by mechanical agitation, it spreads spontaneously and rapidly at the surface of the film, virtually sweeping the film away as it does so. Liquids frequently used for this purpose are octyl alcohol (2-ethylhexanol), tributyl phosphate, silicone (polysiloxane), and many patented preparations. See Adsorption; Interface of phases; Surface tension. [S.RO.]

Bibliography: J. J. Bikerman, *Foams: Theory and Industrial Applications*, 1953; S. Ross, *The Inhibition of Foaming*, Rensselaer Polytech. Inst. Eng. and Sci. Ser. 63, 1950.

Focal length

A measure of the collecting or diverging power of a lens or an optical system. Focal length, usually designated f' in formulas, is measured by the distance of the focal point (the point where the image of a parallel entering bundle of light rays is formed) from the lens, or more exactly by the distance from the nodal point to the focal point. See Optics, geometrical.

The power of a lens system is equal to n'/f', where n' is the refractive index in the image space (n' is usually equal to unity). A lens of zero power is said to be afocal. Telescopes are afocal lens systems. See Telescope; see also Diopter; Lens, optical. [M.H.]

Fog

Water droplets or, rarely, ice crystals, suspended in the air so as to reduce visibility appreciably; a cloud resting on the ground. Fog is most common near sea coasts or large lakes, on mountains, in small valleys of nonarid regions, and over oceans where the water is relatively cold, as in summer off Newfoundland or California. On land, fog is most frequent in the early morning.

Related phenomena. Mist, generally known in the United States as light fog, can be differentiated from true fog by its horizontal visibility. This visibility is more than 1000 meters (5/8 mi) according

to internationally accepted definition. Also, mist may be wet haze (damp haze), composed of wet hygroscopic particles such as sea salt which absorb water in moist but not necessarily saturated air. Dry haze is composed of minute dry particles, and may include smoke or dust, though these are classified separately when possible. Fog, mist, and wet haze are hydrometeors, a general term for all atmospheric water droplets or ice particles. A combination of fog with smoke (and often other pollutants) is called smog.

Low stratus clouds may develop when fog is lifted by turbulence or by warming of the ground; in turbulent conditions low stratus clouds may form initially instead of fog.

Formation of fog. Fog results either when air is cooled to its dew point and below, or by an increase of moisture through evaporation from water that is warmer than the air. Causes of cooling are: the especially rapid nighttime radiation that develops with clear skies; advection (air moving over a colder surface); and expansional (adiabatic) cooling, as when air moves up a mountain slope. For radiation to cause fog, winds must be light or calm; otherwise, mixing prevents fog or produces low clouds. Rapid evaporation into cold air moving over warmer water produces, for example, arctic sea smoke, a thin, wispy fog. Evaporation from warmer raindrops falling through colder air often causes fog.

Before fog can form, there must be sufficient condensation nuclei in the air. This condition is usually met, but ice nuclei are often lacking in sufficient quantity so that many fogs below freezing are water fogs. In the western plateau of the United States supercooled water fog, known locally as pogonip, sometimes forms over a snow cover with temperatures a few degrees below freezing; it leaves a heavy deposit of rime on all outside objects. Fog with snow cover, however, is common only near freezing, and again, at temperatures below about −40°C (−40°F). *See* CLOUD; CLOUD PHYSICS; DEW; DEW POINT; HUMIDITY; PRECIPITATION (METEOROLOGY); SMOG; VAPOR PRESSURE. [J.R.F.]

Bibliography: T. F. Malone (ed.), *Compendium of Meteorology*, 1951; S. Petterssen, Weather Analysis and Forecasting, vol. 2, 2d ed., *Weather and Weather Systems*, 1956.

Fold and fold systems

Folds are recognized where layered rocks have been distorted into wavelike forms. Some folds are fractions of an inch across and have lengths measured in inches, whereas others are a few miles wide and tens of miles long.

Elements of folds. The axial surface (Fig. 1) divides a fold into two symmetrical parts, and the intersection of the axial surface with any bed is an axis. In general, an axis is undulatory, its height changing along the trend of the fold. Relatively high points on an axis are culminations; low points are depressions. The plunge of a fold is the angle

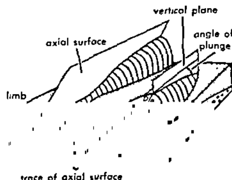


Fig. 1. Elements of folds.

described in terms of the inclination of its limbs and axial surface (axial plane). A symmetrical fold is one in which the axial surface is essentially

horizontal (Fig. 2d). In an isoclinal fold the limbs slope in the same direction at equal angles; the axial surface may have any orientation.

Minor folds and cleavage. Most folds are accompanied by cleavage and smaller folds, and the eval-

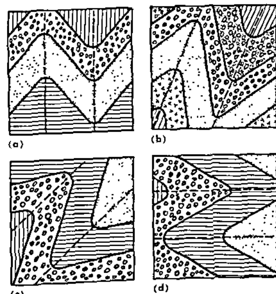


Fig. 2. Geometry of folds. (a) Symmetrical folds. (b) Asymmetrical folds. (c) Overturned folds. (d) Recumbent folds.

uation of the geometric relations between these features has been a major concern of structural geologists.

Minor folds are usually asymmetrical and have orientations which indicate that, where stratigraphic information is available, the younger beds in a folded sequence slide away from the synclinal axes relative to the older beds. Empirical data also indicate that the axes of the minor folds are oriented parallel to the axes of the larger folds. Generalizations of these relations in regions where the structures are not overly complex allow the reconstruction of large folds from scattered observations of minor folds (Fig. 3).

Investigations also show that a geometric relationship exists between cleavage and folding. In general, the cleavage is approximately parallel to the axial surface of a fold or is slightly fanned with respect to this surface. In either case, the acute angle between cleavage and bedding points in the direction in which the adjacent beds are assumed to have slipped, and the intersection of cleavage and bedding is approximately parallel to the axis of the major fold. Assuming that the younger beds slip away from synclinal axes relative to older beds, the cleavage-bedding relations provide a means for reconstruction of the major fold (Fig. 3).

Behavior of rocks with depth. Folds in which the beds have constant thickness are called concentric folds. In these the anticlines become sharper and the synclines broader at depth, until a horizon is

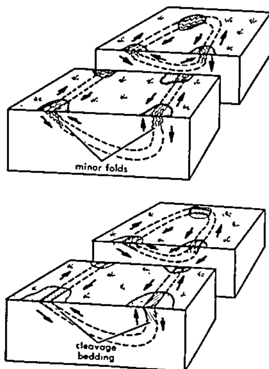


Fig. 3 Use of minor folds and cleavage-bedding relations in reconstructing a major syncline.

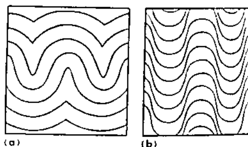


Fig. 4. Types of folding. (a) Concentric folding (b) Similar folding. (*From M. P. Billings, Structural Geology, 2d ed., Prentice-Hall, 1954*)



Fig. 5. Décollement of the Jura Mountains. The lowest formation, with nearly vertical structure, consists of Paleozoic crystalline rocks. Directly above these is a thin bed of flat-lying Triassic quartzite. The lower, solid black formation, which is very incompetent, consists of anhydrite, shale, and salt. The higher beds are Triassic, Jurassic, Cretaceous, and Tertiary sedimentary rocks. (*After A. Buxtorf, from M. P. Billings, Structural Geology, 2d ed., Prentice-Hall, 1954*)

reached below which the fold does not exist (Fig 4*a*). Similar folds are those in which the beds are thinned in the limbs and thickened in the crests and troughs. Folds of this type could theoretically continue downward without a change in form for considerable depth (Fig. 4*b*).

However, where unusually good data are available, most folds are found not to be persistent with depth; that is, most folds are disharmonic. In these a layer exists at some depth which accommodates itself to the folded rocks above and locally acts as a detachment horizon, permitting the upper layers to fold independently of the lower beds. Under certain conditions, the rocks above a detachment horizon are folded independently of the rocks beneath [illegible] off is a

[illegible] sult from differential vertical forces or from compressive forces acting in directions more or less parallel to layering in rocks. In folds resulting from differential vertical forces, a segment of a bed is lengthened by stretching and thinning (Fig. 6*a*). In folds resulting from compressive forces acting in directions parallel to the layering in rocks, a segment of a bed retains its original length, but after folding, its extremities become less widely separated (Fig. 6*b*). The convex side of a bed is subjected to tension, the concave side is compressed, and a surface of no strain separates the two regions (Fig 7). With increasing curvature of the fold, the surface of no strain migrates toward

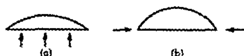


Fig. 6. Deformation of segment of bed in folding. (a) By vertical forces. (b) By horizontal forces. (From M. P. Billings, *Structural Geology*, 2d ed., Prentice-Hall, 1954)

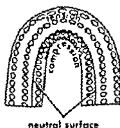


Fig. 7. Bending of a bar of plastic material. Circles inscribed before deformation indicate the state of strain in the various parts of the bar. (After B. Sander, *Gefügekunde der Gesteine*, 1930, from E. S. Hills, *Outlines of Structural Geology*, 3d ed., Wiley, 1953)

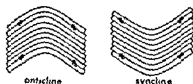


Fig. 8. Cross sections illustrating flexure folding (From M. P. Billings, *Structural Geology*, 2d ed., Prentice-Hall, 1954)

the convex side. This migration is particularly marked in beds of low strength.

Folds involving beds of considerable strength tend to be concentric. Each bed slips past the adjacent bed in such a way that the younger moves away from the synclinal axes and toward the anticlinal axes (Fig. 8). The rarity of deformed fossils in these beds suggests that internal distortion is minimal.

Groups of beds of differing strength tend to be deformed into similar folds. The relatively strong beds retain their original thicknesses and control the folding. The weaker beds are thinned in the limbs and thickened in the crests and troughs, thus accommodating themselves to the form of the stronger beds.

It is not clear how the distortion in the weaker beds develops. In general, the relatively weak beds are cut by cleavage, and this has led some geologists to believe that the distortion took place by slip along cleavage planes and flattening in a direction at a large angle to them. Others believe that the distortion is the result of plastic flow of the weak rock, the cleavage being a later development.

It has been recognized that if all the rocks in a fold are cut by cleavage, the fold may owe its shape to slip along the cleavage and flattening of the plates between. Folds interpreted as having originated by such a mechanism are called slip folds. On the other hand, if the rocks of a fold are interpreted as being deformed principally by plastic flow, the structure is called a flow fold. The present state of knowledge, even where the strain in rocks has been systematically measured, does not allow an unequivocal solution to the problem of folding.

Fold systems. Folds generally do not occur singly but are arranged in festoons in mobile belts with lengths of thousands of miles and widths of hundreds of miles The folds of these belts, or fold systems, commonly consist of great complex structures composed of many smaller folds. In this regard, a large anticlinal structure consisting of numerous smaller folds is called an anticlinorium, and a large synclinal structure consisting of many smaller folds is a synclinorium. The trends of these smaller folds are more or less parallel, although the relations may be complex. In some regions folds belonging to several generations may be present, and locally these may have divergent trends.

Sections across deformed belts indicate that in marginal areas the folds tend to be simple and asymmetrical away from the medial parts of the fold system. These folds are commonly accompanied by thrusts which dip toward the inner portion of the belt. Toward the central part of the fold system the folds become increasingly complex and have forms suggesting that the rocks were highly plastic during the deformation. The interior zones are also marked by more or less complete recrystallization of the rocks and extensive plutonic activity See PLUTON; TECTONIC PATTERNS.

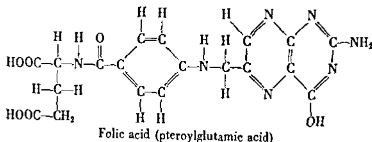
[P.H.O.]

Bibliography: M. P. Billings, *Structural Geology*, 2d ed., 1954; W H. Bucher, *The Deformation of the Earth's Crust*, 1941; L. U de Sitter, *Structural Geology*, 1956; J. Goguel, *Introduction à l'étude mécanique des déformations de l'écorce terrestre*, 1943; E. S. Hills, *Outlines of Structural Geology*, 3d ed., 1953; P. B. King, *The Tectonics of Middle North America*, 1951.

Folic acid

A yellow vitamin, slightly soluble in water, which is usually found in conjugates containing varying numbers of glutamic acid residues. It is also known as pteroylglutamic acid (PGA), and has the structural formula shown on the next page.

It is now known that vitamins M and B_c, factors R, S, and U, the *Lactobacillus casei* factor, and a number of others are pteroylglutamates. Folic acid is usually assayed microbiologically using *L. casei* or *Streptococcus faecalis*. Because of the various forms in which the vitamin appears in nature, some having different activity for the two organisms, and because of the difficulty in hydrolyzing the PGA conjugates, it is extremely difficult to determine absolute amounts of the vitamin. See BIOASSAY.


Folic acid (pteroylglutamic acid)

A formylated derivative of folic acid, called *citrovorum factor, folinic* acid, or leucovorin, has been isolated. It appears that folinic acid or closely related compounds may be the forms in which folic acid is incorporated into coenzymes. Liver and kidney enzymes have been obtained which catalyze the conversion of folic to folinic acid. Ascorbic acid appears to be involved in this conversion. Folic acid enzymes are particularly concerned with the metabolism of methyl groups (CH_3). They are involved in the synthesis of serine, methionine, choline, purines, and pyrimidines. These functions of folic acid involve it in the metabolism of nucleoproteins, which are important in cell division. *See* NUCLEOPROTEIN.

Folic acid is required by a large number of animals and microorganisms. Many animals obtain enough folic acid from intestinal synthesis so that a dietary source is not necessary. In these species, the feeding of antibiotics or sulfa drugs can precipitate deficiency signs. Folic acid deficiency is usually accompanied by poor growth, anemia, and other blood dyscrasias.

Folic acid is so widespread in nature and intestinal synthesis is so great that a folic acid deficiency in humans because of low dietary intake is probably very rare. Deficiencies of other nutrients leading to impaired folic acid metabolism or difficulties in hydrolysis of folic acid conjugates by intestinal enzymes or poor absorption may lead to a number of clinical conditions in which folic acid deficiency is involved. These include various nutritional macrocytic anemias, sprue, idiopathic steatorrhea and pernicious anemia. The daily requirement of humans for folic acid is unknown, but on the basis of food analyses and therapeutic experience, it appears to be less than 1 mg. *See* VITAMIN.

Folic acid antagonists such as aminopterin, which inhibits the conversion of folic to folinic acid, have been used in the treatment of leukemia. Although such treatments have not cured this disease, they have in many cases brought about temporary remissions. [S.N.C.]

Industrial production. Folic acid is produced commercially by the simultaneous reaction of 1,1,3-tribromacetone with 2,5,6-triamino-4-hydroxypyrimidine and *p*-aminobenzoylglutamic acid. Dibrompropionaldehyde, with the assistance of an oxidizing agent, is sometimes used as a [illegible] good yield of crude folic acid. Purification of the crude through the zinc and magnesium salts gives a golden product which is essentially free of impurities.

Various processes can be used to make folic acid, but this one has been given because of its relative simplicity. The intermediates can be made readily by standard techniques familiar to organic chemists. Most of the other processes are based on a stepwise conversion of the pyrimidine to the pteridine ring, along with additions to the benzoylglutamic molecule. Folic acid is then produced in a two-component condensation, rather than a three-component condensation. [R.J.T.]

Fontéchevade man

A fossil man, known from two skull fragments from the Fontéchevade cave, Charente, France, represents Third Interglacial *Homo sapiens*. The fragments exhibit brow ridges of the type known only in modern man, and a cranial vault of modern form. The cave contained the following layers from top to bottom: Upper Paleolithic and Mousterian, of Würm (Last) Glacial date; a stalagmitic layer separating these from all lower contents; and a lower layer containing a Third Interglacial fauna and a pre-Mousterian Tayacian stone industry. The cave was excavated in 1947 by Mlle G. Henri-Martin. *See* FOSSIL MAN. [W.W.H.]

Food

The material that enables man to grow and to maintain and reproduce himself. It is essential to life, as are water and oxygen, which are important constituents of foods.

The growth of modern urban civilizations and an expanding world population have led to the development of new and revolutionary methods in agriculture. Concurrent with these, new and improved methods of processing and distributing foods have resulted in the growth of a vast and important industry. In the United States, this industry amounts to $75,000,000,000 per year.

The food engineer is concerned with the problems of food transfer from the field to the consumer, problems of freshness of foods, attractiveness, and stability, the prevention of spoilage and avoidance of contamination, and the manufacture of new products from food ingredients. Most particularly, he is concerned with the maintenance or enhancement of the nutritional value of the product in the processing procedures. Food additives

which play a large part in these problems and procedures are discussed in the latter portion of this article.

In any dietary regimen minimal amounts of at least three types of materials must be provided: (1) energy-producing compounds, (2) structure-producing compounds and (3) compounds, not of themselves energy producers, necessary to energy-exchanging reactions. These three types of compounds include the proteins, carbohydrates, fats, vitamins, and mineral elements. Water may be considered essential for their utilization.

GENERAL CONSIDERATIONS

The essential food components, staple foods and processed and refined foods, are discussed in this section.

Essential components. Energy for the various life processes is supplied by carbohydrates, fats, proteins, vitamins, and minerals. Energy requirements are determined by the age, sex, and activity of the individual.

Energy requirements. Strenuous activity requires more calories per day than does sedentary work, for example, 4500 calories (cal) as compared to 2400 cal *for a 70-kilogram male. Energy* values in kilocalories per gram have been calculated to be about 4.1 for carbohydrate, 4 1 for protein, and 9.3 for fat.

Carbohydrates. These are generally thought of as sweetening agents but are not limited to this function. Carbohydrates include glucose, fructose, maltose, lactose, sucrose, glycogen, dextrin, starch, and cellulose and are utilized chiefly as sources of heat and energy. Cellulose, a carbohydrate, is utilized chiefly as roughage. *See* CARBOHYDRATE

Proteins. These provide the nitrogenous material necessary for structure and are large molecules, composed of amino acids *joined in peptide* linkage. The proteins function in tissue growth and repair, the formation of antibodies, enzymes, some hormones, hemoglobin, maintenance of acid-base balance, and the formation of hereditary material (*see* PROTEIN). At least eight amino acids are considered essential for adult man (*see* AMINO ACIDS). The following amino acids may be obtained from protein hydrolysates.

Essential or indispensable	*Partially essential (or partially indispensable)*	*Nonessential or dispensable*
Isoleucine	Arginine	Alanine
Leucine	Cysteine	Aspartic acid
Lysine	Cystine	Glutamic acid
Methionine	Histidine	Glycine
Phenylalanine	Tyrosine	Hydroxyproline
Threonine		Proline
Tryptophane		Serine
Valine		

Fats. Fats are esters of fatty acids with various alcohols. A knowledge of the human requirements for fatty acids is still far from complete, but three, linoleic, linolenic, and arachidonic, are considered *essential or indispensable.* "Essential" or "indispensable" implies that the body either cannot synthesize them at all or not rapidly enough for man to be independent of a dietary supply; therefore, the component must be supplied in the diet. *See* LIPID.

Vitamins. Vitamins, originally designated accessory food *factors, are organic compounds* necessary for normal nutrition and health and must be supplied in the diet. Many of them serve as coenzymes (*see* COENZYME). Others have functions not fully clarified, but all serve some function in *cell chemistry. Requirements vary with age, sex,* and occupation. Classification of the vitamins has been on an alphabetical basis, a fat- or water-soluble basis, and more recently, with the establishment of chemical structure, by specific chemical title (Table 1). Absence of a vitamin from the diet leads to deficiency symptoms which may be relieved by administration of adequate amounts of the vitamin in question. The strength of vitamin preparations usually is stated in international units based on the activity of an international standard preparation, milligrams, or micrograms.

Table 1. Known vitamins*

Letter	Name	Function
Fat soluble		
A†	Axerophthol	Important to vision and general health
D_2†	Calciferol	Antirachitic factor
E†	Tocopherol	Antisterility factor
K†	Phylloquinone	Antihemorrhagic factor
Water soluble		
C	Ascorbic acid	Oxidation-reduction systems in cells, antiscorbutic factor
H	Biotin	Probably acts as a coenzyme
	Choline	One of an essential methyl transfer group
B_{12}†	Cyanocobalamine	Antipernicious anemia extrinsic factor
	Inositol	Related to metabolism of glucose
	Niacin	Coenzymes I and II (pellagra preventive)
B_3	Pantothenic acid	Related to acetyl coenzyme A
	p-Aminobenzoic acid	A component of the folic acid group
M, Bc	Folic acid (pteroylglutamic acid)	Necessary for normal functioning of hematopoietic system
B_6†	*Pyridoxine*	Coenzyme metabolism of amino acids and of fats and fatty acids
B_2, G	Riboflavin	Coenzyme systems involved in phosphorylations (respiratory enzymes)
B_1	Thiamine	Involved in oxidative decarboxylations (cocarboxylase)

* See individual articles on each vitamin.
† These vitamins are covered in separate arti[illegible] under vitamin The others are listed by name

Minerals. Small amounts of certain inorganic substances, known as mineral elements, are essential to good health. The requirements for these appear to be associated with the state of vitamin nutrition and with endocrine activity. Included among the needed minerals are the following:

Calcium	Catalyst, enzyme activator, milk production
Phosphorus	Metabolism, acid-base regulation, vitamin and enzyme activity
Magnesium	Protein formation from amino acids
Sodium	Regulator of neutrality; balances basic ions; muscle contraction
Potassium	Muscle function, nerve excitability, carbohydrate metabolism
Manganese	Enzyme activator, affects thiamine utilization; related to reproduction physiology
Iron	Catalyst
Copper	Catalyst
Cobalt	Found in vitamin B_{12}
Iodine	Constituent of thyroxin
Sulfur	Part of the amino acids
Zinc	Carbon dioxide metabolism; functioning of pancreas

Other minerals such as aluminum, bromine, fluorine, and silicon appear to be necessary, but their exact function is not clear

Table 2 lists recommended daily dietary allowances compiled by the Food and Nutrition Board, National Research Council.

STAPLE FOODS

The basis of the dietaries of civilized man are the great staple foods, which may be classified as foods of plant origin and foods of animal origin

Foods of plant origin
- Cereals: wheat, corn (maize), rye, barley, oats, rice
- Sugar: sugar cane, sugar beets
- Vegetables: potatoes, sweet potatoes, cassava, beans, peas, onions, tomatoes
- Fruits: banana, plantain, berries
- Nuts: almonds, coconuts, various seeds
- Bacteria: molds, mushrooms, yeasts

Foods of animal origin:
- Meats: beef, pork, mutton, poultry
- Eggs
- Dairy products
- Fish and fish products

Foods of plant origin. Foods of plant origin are discussed in the following sections.

Cereals. These are universally considered essential. They contain carbohydrates and protein, some mineral salts, and fat. They are easily dried, take up little moisture, and keep readily in the natural state. The following cereals are of particular interest:

Wheat is an ancient and highly respected food material; it grows in regions of relatively low rainfall and wide extremes of temperature. A number of varieties are grown, depending on cultivation in winter or spring. Crops are subject to rusts and damage from insects. Wheat yields flours of different grades, and its largest use is in the manufacture of bread and other baked products. *See* WHEAT

Corn, the most valuable cereal crop of the American continent, is also known as maize. Corn is used as "green corn," grain as raw material for

Table 2. Daily dietary allowances[1]

	Age years	Weight kg (lb)	Height cm (in.)	Calories	Protein, g	Calcium, g	Iron, mg	Vitamin A, IU	Thiamine, mg	Riboflavin, mg	Niacin,[2] mg equiv	Ascorbic acid, mg	Vitamin D, IU
Men	25	70 (154)	175 (69)	3200[3]	70	0.8	10	5000	1.6	1.8	21	75	
	45	70 (154)	175 (69)	3000	70	0.8	10	5000	1.5	1.8	20	75	
	65	70 (154)	175 (69)	2550	70	0.8	10	5000	1.3	1.8	18	75	
Women	25	58 (128)	163 (64)	2300	58	0.8	12	5000	1.2	1.5	17	70	
	45	58 (128)	163 (64)	2200	58	0.8	12	5000	1.1	1.5	17	70	
	65	58 (128)	163 (64)	1800	58	0.8	12	5000	1.0	1.5	17	70	
	Pregnant (second half)			+300	+20	1.5	15	6000	1.3	2.0	+3	100	400
	Lactating (850 ml daily)			+1000	+40	2.0	15	8000	1.7	2.5	+2	150	400
Infants[4]	0–1/12[4]				See Footnote 4								
	2/12–6/12	6 (13)	60 (24)	kg × 120		0.6	5	1500	0.4	0.5	6	30	400
	7/12–12/12	9 (20)	70 (28)	kg × 100		0.8	7	1500	0.5	0.8	7	30	400
Children	1–3	12 (27)	87 (34)	1300	40	1.0	7	2000	0.7	1.0	8	35	400
	4–6	18 (40)	109 (43)	1700	50	1.0	8	2500	0.9	1.3	11	50	400
	7–9	27 (60)	129 (51)	2100	60	1.0	10	3500	1.1	1.5	14	60	400
	10–12	36 (79)	144 (57)	2500	70	1.2	12	4500	1.3	1.8	17	75	400
Boys	13–15	49 (108)	163 (64)	3100	85	1.4	15	5000	1.6	2.1	21	90	400
	16–19	63 (139)	175 (69)	3600	100	1.4	15	5000	1.8	2.5	25	100	400
Girls	13–15	49 (108)	160 (63)	2600	80	1.3	15	5000	1.3	2.0	17	80	400
	16–19	54 (120)	162 (64)	2400	75	1.3	15	5000	1.2	1.9	16	80	400

[1] The allowance levels are intended to cover individual variations among most normal persons as they live in the United States under usual environmental stresses

[2] Niacin equivalents include dietary sources of the preformed vitamin and the precursor, tryptophan 60 milligrams tryptophan equals 1 milligram niacin

[3] Calorie allowances apply to individuals usually engaged in moderate physical activity Adjustments must be made for variations in body size, age physical activity, and environmental temperature

[4] The Board recognizes that human milk is the natural food for infants and feels that breast feeding is the best and desired procedure for meeting nutrient requirements in the first months of life No allowances are stated for the first month of life Recommendations as listed pertain to nutrient intake as afforded by cow's milk formulas and supplementary foods given the infant when breast feeding is terminated Allowances are not given for protein during infancy

Source *Recommended Dietary Allowances*, rev ed., National Academy of Sciences—National Research Council, Publ 589 1958

many other products as corn meal, hominy, and cornstarch, and in the manufacture of breakfast foods such as corn flakes and puffed corn. It is also used in fermentations as basic raw material for grain alcohol and distilled beverages, butyl alcohol and acetone, and as a raw material for manufacture of sugar in the form of glucose, syrups, and other products. The greatest use of corn is as a cattle feed. *See* CORN; DISTILLED SPIRITS; ETHYL ALCOHOL.

Rye is an extremely hardy plant. It is used as a grain crop for animal feed, a raw material for the manufacture of distilled liquors, particularly rye whiskey, and as a component of certain types of bread. *See* RYE.

Barley is a cereal that is a food for man and animals and a source of malt for brewing. Barley malt is used for breakfast foods and special food preparations; for malt syrups which may find use in baking industries; for flavoring foods or for medicines; in brewing beverages; as a diastatic or desizing agent in textile processing. *See* BARLEY.

Oats is an excellent food for animals; it is eaten by humans in the form of oatmeal and rolled oats or breakfast foods. *See* OATS.

Rice grows in many varieties and is a nutritious and easily digested food. Rice in its many varieties serves as the main source of food for one-third of the world's population. *See* RICE.

Milling includes the various operations involved in the transformation of cereals, particularly wheat, into flour. These are largely mechanical operations (grinding, reduction, bleaching by chemical means, grading) and depend on a reduction in size of the components which make up the wheat kernel, with a subsequent separation of the particles according to size and composition. This is a food processing operation.

Sugars and syrups. The sugar cane is crushed, the juice clarified, the raw sugar crystallized and separated from the molasses, and further refined by a series of treatments. Beet sugar processing differs somewhat in that the sugars are removed initially by repeated diffusion with warm water. Syrups, of which cane and maple are representative, are the direct result of concentrating all the sugar of the original sap. *See* SUGAR.

Sugar industries of the United States constitute an important part of the food industries. Large quantities of raw sugar are imported to be refined to supplement our native supply. Sugar beets grow best in temperate climates and are an important source of sugar. Beet pulp residues, and tops and crowns represent a valuable stock fodder.

Vegetables. These include a great variety of plant products. Some are eaten raw as greens and salad materials, whereas others are boiled, baked, or cooked in conjunction with meats or other food materials.

The more extensive utilization of vegetables in the diet has been partly responsible for the development and improvement of preservation processes. The cultivation, harvesting, packing, storage, and shipment involve many problems. Among these are soil, climate and growth, physiological and microbiological damage, plant diseases, bruising or injuring during handling, ventilation, and storage temperatures. Potatoes, utilized also as flour and starch, are the most valuable vegetable crop raised in the world and are used for human food, stock feed, fermentations, and for industrial ethyl alcohol. Next to potatoes, tomatoes rank highest as a vegetable crop. They are high in water content, and contain malic acid, sugar, and heat-stable vitamins. Carrots, beans, and onions are also valuable crops.

Fruits. Fruits are a source of sugar, organic acids, and mineral salts. They are uniformly high in moisture content and low in starch.

Apples, known and appreciated as a food for many centuries, are used as fresh fruit, for cooking in various ways, and for canning. Apple juice is a source of cider and cider vinegar. Pectin is an apple by-product used in the manufacture of jellies.

Principal citrus fruits of the United States are oranges, grapefruit, and lemons. Limes, tangerines, and kumquats are also of commercial interest. Citrus fruits are used as fresh fruits, fresh and frozen juices, and as a source of citric acid and pectin. See articles on individual citrus fruits.

Other fruits of commercial importance include strawberries, pears, grapes, peaches, bananas, and cranberries. See articles on the individual fruits.

Nuts. These are almonds, coconuts, and various seeds. The nuts and seeds are high in fat and protein content and contain fair amounts of carbohydrates.

Microbes. Molds, yeasts, and mushrooms have increased in importance as foods. Yeasts have been used since primitive times, but newer methods of culturing these and mushrooms have extended their marketability. *See* MUSHROOM; YEAST, INDUSTRIAL.

Foods of animal origin. These are predominantly protein, very low in carbohydrate content, and contain varying amounts of fat. Of the many animal foods, milk most nearly approaches the per-

[illegible]

and pork [illegible] products include lard from hog fat, stearin and tallow from beef fat, sausage, made of a number of meat products, and gelatin from hides and trimmings.

Poultry includes domestic birds, alive or dressed, which have been bred and raised for eating purposes, egg production, or for show or exhibition purposes. These birds include chickens, domestic fowl, ducks, geese, turkeys, and pigeons. Birds are killed by bleeding, feathers and the contents of the body cavity are removed, and the bird is immediately chilled by one of several methods. *See* POULTRY PRODUCTION.

Eggs. With proper humidity, adequate ventilation, and freedom from odors, eggs may be stored,

successfully for periods of 9 months or longer if kept at a temperature slightly above their freezing point. Eggs may be cleaned by sand-blasting or washing. Egg contents may be frozen or dried. *See* EGG PROCESSING.

Dairy products. Milk is usually pasteurized, that is, subjected to a temperature of 143°F (61.7°C) for at least 30 min, or 161°F (71.7°C) for 15 seconds and then cooled and bottled. The importance of safety and cleanliness is stressed in the dairy industry. Milk may also be condensed or evaporated, dried, powdered, or separated into skim milk and fat. *See* MILK.

Butter is cream, subjected to churning operations, salted, packed, and stored. Margarines are hydrogenated fats with flavors similar to butter. *See* BUTTER.

Ice cream is the frozen product made from a combination of milk products (cream, butter, or milk—evaporated, condensed, skimmed, or dried) and two or more of the following ingredients: eggs, water, and sugar, with harmless flavoring and harmless coloring matter, and with or without stabilizer. In the manufacture of ice cream, freezing is accompanied by agitation of the ingredients. *See* ICE CREAM.

Cheese is defined as the product made from curd obtained from the whole, partly skimmed, or skimmed milk of cows, or from the milk of other animals, with or without added cream, by coagulating the casein with rennet, lactic acid, or other suitable enzyme or acid. The separated curd is used with or without further treatment by heat or pressure, or by means of ripening ferments, special molds, or seasoning. *See* CHEESE.

Fish and fish products. Fish and fish products are cooled by ice or refrigeration until port is reached. Further preservation methods for longer periods of time include freezing, canning, salting, smoking, and drying. Shellfish are subjected to special chlorine treatment.

Processed and refined foods. Processed foods are foods intentionally subjected to special treatment before consumer purchase in order to extend storage life. A refined food may be thought of, not as a complete food, but rather an ingredient or supplement in the diet or in manufactured foods. The distinction between processed and manufactured foods is not sharp; for example, canned whole kernel corn may be thought of as a processed food, while cornflakes might be considered to be a manufactured food. Food processing has been made possible by food engineering unit operations and unit processes; by bringing together the essential knowledge regarding the chemistry, microbiology, and nutritional qualities of foods, and the engineering skill to apply and control the various operations involved in all the manufactural treatments which eventually yield the commercial foods of today. In the processing, contamination by biological or chemical agents (microorganisms, insects, pests, trace metals, or container odors) must be avoided, and care is taken to preserve nutrient materials in the food and to prevent changes in palatability attributed to changes in odor, flavor, and texture, and other organoleptic characteristics. Consumer acceptability of a quality product is the final test of the processing treatment. Foods and varieties of the same food differ in their suitabilities for processing. A process for one food may be quite unsatisfactory as a process for another food. Methods of processing foods include: (1) drying, or dehydration; (2) salting, or use of other preservatives; (3) freezing; (4) freeze-drying; (5) heating, as in pasteurization and sterilization; (6) ionizing radiation; and (7) treating with antibiotics.

Concurrent with the development of processing methods has been the growth of the packaging (tin cans, glass containers, flexible wrappings) and refrigeration industries.

Food spoilage. All foods are subject to spoilage and the extent of spoilage will depend upon a number of factors, such as number and kind of microorganisms present, the number and kind of enzymes present, whether the food has been pretreated or not by sulfur dioxide, germicidal vapors, or rays of various kinds. Environmental factors will affect the keeping qualities. Such factors include the temperature and humidity at which the food is held; whether the food has been frozen, heated, moistened, or dried; and the particular nature of the food itself, its composition or structure. A food that is thought fit to eat by one person may be quite unattractive to another. In general, fitness of food is determined by its state of development or maturity, absence of contamination in processing, and presence or absence of microbial or enzymatic action. Causes of spoilage may be microorganisms, insects and other pests, enzyme activity, natural metabolic reactions, and environmental effects such as freezing, burning, drying, and so forth.

Spoilage classification. A food spoilage classification may be set up as follows:

1. Staple or nonperishable foods do not spoil unless handled carelessly and include sugar, flour and cereal grains, and dry beans.

2. Semiperishable foods, which remain unspoiled for a fairly long period with proper care, include potatoes, certain varieties of apples, nut meats, onions, dried fruits, and other dried foods.

3. Perishable foods, which spoil readily without special preservative methods, include meats, fish, poultry, most fruits and vegetables, eggs, milk and milk products, and shellfish.

Food spoilage may be considered as any change in the natural state of the food that lessens its desirability, either for aesthetic or health reasons. In general, these changes are of a biological nature, brought about by the growth of microorganisms, or are autolytic changes brought about by enzymes produced within the tissue itself. In some cases, changes result from oxidation. These factors may act alone or in combination. Microorganisms may synthesize compounds that alter the flavor, color, or odor of the food.

Microbial decomposition products. Products that may be formed because of the microbial decomposition of carbohydrates, fat, and protein, may be listed as follows:

Carbohydrates	*Fats*	*Proteins*
Alcohols	Alcohols	Aldehydes
Aldehydes	Aldehydes	Aliphatic acids
Aliphatic acids	Aliphatic acids	Amines
Carbon dioxide	Carbon dioxide	Amino acids
Hydrogen	Fatty acids	Ammonia
Ketones	Glycerol	Aromatic acids
Methane	Hydrogen	Heterocyclic compounds containing indole and other phenolic groupings
	Ketones	Hydrogen
	Methane	Carbon dioxide
		Polypeptides
		Proteoses
		Peptones

Decomposition of proteins results in the splitting of the protein molecule into constituent groups and is usually associated with the liberation of odoriferous compounds.

Although fats are less susceptible to microbial decomposition, many microorganisms contain lipolytic enzymes which hydrolize the fat to fatty acids and glycerol and may also produce rancidification. The breakdown of the fatty acid portions of the fat may occur with the formation of shorter chained acids, aldehydes, ketones, and peroxides.

Microbial activity affecting carbohydrates results in acid and gas production known as fermentation. If air has free access to the food, such acids will not be stable. The higher carbohydrates may undergo some change due to microbial action, for example, the conversion of starch sugars and acids.

Spoilage causes. Particular food products and causes of spoilage are given below:

In the case of bakery products, grain products, and meals and flours made from these products the moisture content determines the amount of microbial spoilage. With a slight moisture content molds grow, and at an increased moisture content microbial growth takes place. If properly prepared and packaged, no growth occurs.

In sugars and sugar products, spoilage is limited to those microorganisms that are able to grow in concentrated solutions of the sugar. This includes certain yeasts and molds.

Vegetables and fruits are spoiled chiefly by plant pathogenic microorganisms and rots of various kinds.

Raw meat may be spoiled by microbial activity, surface slime and molds, enzymatic autolysis, and fat oxidation. The physiological state of the animal prior to killing, the method of killing and bleeding, and the rate of cooling will determine, to a large degree, how extensively microorganisms will develop and spoil the meat.

Fish is spoiled chiefly by microorganism growth but also by autolysis and oxidation when processed and stored.

Undesirable changes in egg, poultry and milk are brought about by the growth of microorganisms. Bacteria are common agents of spoilage, unless refrigeration is carefully utilized.

The prevention, or retardation, of these spoilage effects is a primary aim of the food processing industries.

Nonmicrobial food spoilage. Many foods naturally contain enzymes, not of microbial origin, that produce changes in the food under proper conditions. The moisture content of the food also affects its keeping qualities. A blanched, dried product will keep well, whereas unblanched products will spoil because of enzyme activity. Molds will grow on some products of relatively low moisture content, such as dried fruits. Prevention of spoilage is discussed in the previous section. [B.E.P.]

FOOD ADDITIVES

To supplement the ancient methods of food preservation such as salting, drying, smoking, and sulfuring, a wide variety of substances known as food additives has come into use in modern agricultural and food technology. Their uses have extended beyond mere preservation to include all stages of production, processing, and distribution of food. Among the major purposes of food additives are the retention or improvement of nutritive value, color, flavor, appearance, or texture; the prevention of deterioration or spoilage, whether of microbiological or chemical origin during handling, storage, and transportation; the facilitation of mass production of products of uniform quality

home.

Comprehensive tabulations of food additives in common use are contained in Publication No. 398

term, food additives are chemical substances intentionally introduced into food to impart a desired quality or to facilitate production. Insecticides, fungicides, rodenticides, and other pesticidal agents used in agricultural technology may remain in foods as trace residues, but these are regarded as incidental or unintentional residues rather than as food additives and are subject to separate and special regulatory control under the Insecticide, Fungicide, and Rodenticide Act, and the Food, Drug, and Cosmetic Act.

Functional classification. Table 3 presents the principal functional classes, types, and uses o

Table 3. Functional classification of food additives

Class	Principal function	Examples	Used in
Acidulants	To impart acidity or to compensate for deficiency in natural acidity	Acetic, citric, malic, phosphoric acids	Jellies, jams, preserves, soft drinks, confections
Aerating agents	To charge with gas or to displace air	a. Carbon dioxide b. Nitrogen c. Nitrous oxide	a. Carbonated beverages b. Oxidizable oils, milk powder c. Pressure-dispensed creams, etc.
Anticaking agents and desiccants	To prevent lumping or caking	Calcium phosphate Magnesium carbonate, calcium or magnesium stearate Silica gel, talc, starch	Table salt, confectioners' sugar Malted milk powder, onion or garlic salt, meat-curing mixes
Antioxidants	To retard oxidative deterioration, rancidity in fats and oils or off-flavor in frozen foods, flavoring oils, beverages, etc.	a. Sulfur dioxide and sulfites b. Ascorbic and isoascorbic acid and their sodium salts; ascorbyl palmitate c. Nordihydroguaiaretic acid; propyl gallate; butylated hydroxytoluene and butylated hydroxyanisole d. Lecithin, tocopherols	a. Wine and beer; sugar sirups; cut, peeled, or dried fruits and vegetables b, d. Cooked, cured, comminuted meat products, frozen fish dip, beer; flavoring oils and emulsions; confectionery c, d. Lard, vegetable oils, hydrogenated shortenings; crackers, biscuits, breakfast cereals; dry cake or soup mixes
Bleaching agents	To remove or lighten (especially carotenoid) color of certain natural products	a. Oxides of nitrogen, chlorine, chlorine dioxide, nitrosyl chloride b. Benzoyl peroxide c. Ammonium persulfate	a, b, c. Wheat flour b. Cheese (blue or gorgonzola) made from cow's milk
Buffering agents	To adjust and stabilize pH	a. Sodium acetate, citrate, phosphates, pyrophosphates b. Acetic, citric, lactic acids	a. Jellies, jams, preserves; soft drinks; confectionery Prepared cereals and farinaceous products, meat cures, evaporated milk b. Processed cheese, pickles and brines, canned vegetables, confectionery
Clarifying agents	To remove or prevent turbidity	Pectinases, proteinases (papain, bromelin, fungal enzyme preparations) Tannin, gelatin, albumin, methylcellulose	a. Fruit juices, vinegar, wine, beer, soft drinks
Clouding agents	To impart turbidity and prevent "ringing"	Brominated vegetable oils	"Cloudy" citrus-flavored drinks
Coating agents	To impart surface glaze, to facilitate removal of baked products from pans or confectionery from slabs	a. Shellac, beeswax, paraffin, gum benzoin b. Carnauba wax c. Silicone, mineral oil	a, b. Chocolates and confectionery b. Oranges c. Coating baking pans or as "slab dressing"
Coloring agents	To impart color or to adjust for natural variations in color	Annatto, caramel, carmine, carotene, saffron, turmeric Certified food colors (FD&C "coal-tar colors") Carbon black, titanium oxide	Soft drinks, confectionery, icings, jellies, baked goods, fillings, cake mixes, cereals, popcorn, packed fruits, margarine, cheese, butter Meat products
Color stabilizers	To prevent fading, "browning" or other discoloration	a. Sulfur dioxide and sulfites b. Sodium nitrite and nitrate c. Ascorbic and isoascorbic acids d. Citric acid	a, c. Cut, peeled, dried, or processed fruit and vegetables b, c. Cured meat products (to convert myoglobin to nitric oxide myoglobin) c, d. Heat-processed (canned) vegetables

Table 3. Functional classification of food additives (cont.)

Class	Principal function	Examples	Used in
Conditioners (yeast foods)	To improve elasticity and stability of doughs and facilitate mechanical baking of products of uniform quality	Mixtures of potassium bromate and iodate; calcium peroxide; ammonium chloride, sulfate and phosphates; calcium carbonate, lactate, sulfate, and phosphates; sodium chloride	Bread, rolls, cakes, and other yeast-fermented baked goods
Emulsifying agents (surfactants)	To facilitate and stabilize oil-in-water (or water-in-oil) dispersions	Mono- and diglycerides of fatty acids, acetyl tartaric, sulfoacetate and monosodium phosphate derivatives of mono- and diglycerides; glyceryl lactopalmitate, sorbitan, polyoxyethylene, and polyoxyethylene sorbitan derivatives of fatty acids	Lard, margarine, shortenings; ice cream, frozen desserts; chocolate, confectionery; bread, baked goods, cake mixes; flavor emulsions and concentrates; vitamin concentrates
Enzymes	To hydrolyze proteins, carbohydrates, hemicelluloses, etc.	a. Proteases of plant (papain, bromelin, ficin), animal (pepsin, trypsin, pancreatin), or microbiological (fungal or bacterial) origin b. Amylase, invertase c. Pectinase d. *Glucose oxidase* e. Carotenase	a. Meat tenderization a, c. Clarification of fruit juices and beverages b. Manufacture of invert sugar, preparation of cocoa sirups d. Prevention of "browning" (sugar-amino acid reaction) e. Bleaching flour
Flavoring agents	To impart or enhance flavor or aroma	a. Natural extracts, essential oils, and oleoresins b. Alcohols, aldehydes, ketones, ethers, and esters c. Monosodium glutamate, pyroligneous acid	a, b. Soft drinks, confectionery, chewing gum, gelatin desserts, ice cream, frozen desserts, baked goods, icings c Meat products and sauces
Foam regulators	To stabilize or depress foam	a. Dextrin, peptones, cellulose gums b. Silicone (dimethyl polysiloxane)	a. Beer, root beer, and other beverages which produce a head when dispensed b. To prevent foaming during boiling of sirups and jellies, fermentation of wine, or bottling of fruit juices, beverages, etc.
Growth stimulators	To promote more rapid growth of poultry and livestock	Antibiotics (chlor- and oxytetracyclines), stilbesterol, phenylarsonic acid	Broiler and turkey rations, cattle and hog feed
Humectants	To absorb or retain moisture	Glycerol, propylene glycol, sorbitol	Dried fruits, confectionery, shredded coconut
Hydrolytic agents	To effect hydrolytic cleavage of proteins, carbohydrates, and cellulosic material	Hydrochloric acid, potassium acid tartrate	Manufacture of gelatin, protein hydrolyzates, sodium glutamate, invert sugar, corn sirup, etc.
Leavening agents	To render batters or doughs light and porous, as by fermentation	Sodium bicarbonate (and to a lesser extent calcium or ammonium carbonates) plus an acid (alum, monocalcium phosphate, sodium acid pyrophosphate, potassium acid tartrate), or acid-producing compound (glucono-δ-lactone)	Self-rising flour, cake mixes, cakes, biscuits, crackers, etc.
Maturing agents	To "age" (oxidize) wheat flour so as to yield more elastic, stable, and uniform doughs	Chlorine dioxide, chlorine, and nitrosyl chloride; potassium bromate	Wheat flour, bromated flour

Table 3. Functional classification of food additives (cont.)

Class	Principal function	Examples	Used in
Neutralizers	To correct excessive natural acidity or to neutralize acids used in processing	Sodium, calcium, or magnesium hydroxides or carbonates; calcium or magnesium oxide	Sour cream, butter, and other dairy products; canned olives, tomato soup Wines
Nutrients	To restore nutritional values lost in processing or to supplement natural content of nutrients	Mineral nutrients: reduced iron and iron compounds (sulfate, phosphates); calcium carbonate and phosphates; potassium iodide; sodium fluoride and silicofluoride	Enriched flour, bread, and cereal products; vitamin D milk and other fortified dairy products; iodized salt, fluoridated water supplies
		Vitamins: vitamin A acetate and palmitate, β-carotene, thiamine, riboflavin, niacin, ergocalciferol (vitamin D_2), cholecalciferol (vitamin D_3), pyridoxine hydrochloride, calcium pantothenate, α-tocopherol	Special dietary foods and supplements; animal feeds and concentrates
		Amino acids: lysine, methionine	
Preservatives	To prevent infection or inhibit spoilage due to bacteria, yeast, molds, or other microorganisms	a. Chlorine, hypochlorites b. Benzoic acid and sodium benzoate c. Sorbic acid and sorbates d. Calcium propionate, sodium diacetate e. Chlor- and oxytetracycline	a. Water supplies b. Fruit jellies, jams, preserves, pickles, relishes, margarine, soft drinks c. Cheese wraps, pie fillings, fruit syrups d. Bread ("rope" inhibitors) e. Dressed poultry, frozen fish
Sequestrants	To "trap" trace metals (especially iron and copper) which adversely affect color, clarity, flavor, or stability	a. Citric acid and citrates, including monoisopropyl or stearyl citrate b. Phosphoric acid and phosphates c. Calcium and sodium salts of ethylenediamine tetraacetic acid	a. Frozen fruit, soup base, margarine oil b. Processed cheese, evaporated milk, rendered fats and oils c. Wines, vinegar, soft drinks
Sweetening agents (artificial)	To impart sweetness in place of sugars	Saccharin Cyclohexylsulfamates (calcium or sodium salts)	Low-calorie or "diabetic" foods, such as beverages, canned fruits, desserts, etc
Texturizing agents	To impart firmness or improve texture in processed foods	a. Alums b Calcium chloride, citrate, carbonate, phosphate (monobasic), or sulfate c. Magnesium chloride	a. Pickled cucumbers b. Processed cheese, canned potatoes, or tomatoes c. Canned peas
Thickening agents	To impart body or thickness to fluids or emulsions	Glycerol, sorbitol, gelatin Pectin, vegetable gums (alginates, carragheen, karaya) Cellulose gums (carboxymethyl cellulose)	Processed cheese, salad dressings; ice cream and frozen desserts Chocolate-flavored milk and syrups Confectionery and icings, low-calorie (nonsugar) beverages and canned fruits Pressure-dispensed creams

food additives. Those listed in the third column represent important examples within each category, but they by no means exhaust the list of substances in use. Moreover, it is not intended to imply that each substance mentioned is applicable or employed for every purpose cited in the last column, nor that every usage has official sanction in the United States. Not included in this tabulation are postharvest pesticides which may leave residues but are not, in the legal sense, food additives, or chemicals which may migrate into foods from packaging materials such as plastic films, can enamels, and others.

Legal definition. The Food, Drug, and Cosmetic Act, as amended September 6, 1958, defines a food additive as "any substance the intended use of

which results or may reasonably be expected to result, directly or indirectly, in its becoming a component or otherwise affecting the characteristics of any food (including any substance intended for use in producing, manufacturing, packing, processing, preparing, treating, packaging, transporting or holding food; and including any source of radiation intended for any such use), if such substance is not generally recognized, among experts qualified by scientific training and experience to evaluate its safety, as having been adequately shown through scientific procedures (or, in the case of a substance used in food prior to January 1, 1958, through either scientific procedures or experience based on common use in food) to be safe under the conditions of its intended use; except that such term does not include—

(1) a pesticide chemical in or on a raw agricultural commodity; or

(2) a pesticide chemical to the extent that it is intended for use or is used in the production, storage, or transportation of any raw agricultural commodity; or

(3) any substance used in accordance with a sanction or approval granted prior to the enactment of this paragraph pursuant to this Act, the Poultry Products Inspection Act or the Meat Inspection Act of March 4, 1907, as amended and extended."

Any substance used in accordance with this definition would, broadly speaking, be regarded as a food additive whether deemed safe or not. Though they are often used to protect or stabilize the nutritive properties of foods, additives generally have no intrinsic nutritional value themselves, except for the class of nutrients, such as vitamins, amino acids, and mineral salts, used for fortification or enriching purposes. Processing or transient chemicals, such as caustic alkalies used in oil refining or to facilitate peeling of fruits and vegetables, and subsequently removed by washing, or diatomaceous earths or carbon used for decolorizing or filtering purposes, or solvents volatilized during processing, all of which leave no residual components, are not considered to be food additives. To the extent that the administration of drugs or growth stimulants (such as antibiotics and hormones) to livestock or poultry may, intentionally or otherwise, leave residues in meat, milk, or eggs, they are regarded as food additives. No fundamental distinction can be made between food additives of natural origin and those produced synthetically.

Under the Food Additives Amendment of 1958, all additives in use prior to January 1, 1958, and not deemed, by qualified experts, to be safe on the basis of either appropriate scientific tests or experience in common use in foods, are subject to approval (or exemption) by the Food and Drug Administration. Petitions for a regulation governing the conditions under which a food additive may be used must contain full information concerning its identity, composition, proposed use and effect, analytical determination, and toxicity. Unless exempted from regulatory control, the use of an additive is subject to a regulation which will prescribe the food or classes of foods in which it may be safely used; a specified maximum limit; the manner of addition or use; directions, labeling, packaging requirements; and other conditions of use. *See* Food engineering.

[B.L.O.]

Bibliography: M. B. Jacobs, (ed.), *The Chemistry and Technology of Food and Food Products*, 3 vols., 2d ed., 1951; S. C. Prescott and B. E. Proctor, *Food Technology*, 1937.

Food chain

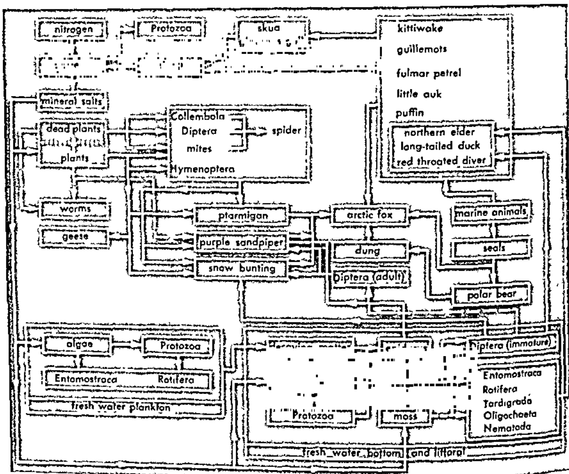
The scheme of feeding relationships which unites *the member species of a biological community*. The idea was introduced to modern ecology by C. Elton to describe the linear series of species, usually involving plants (autotrophs), herbivores, and one or two successive sets of predators (a predator chain), or alternatively, the series of parasites and hyperparasites exploiting a host (a parasite chain). Saprophytic chains, exploiting dead tissues, are now known to be most important. The successive categories in such a chain (see illustration) are widely known as trophic levels.

The number of stages in a given chain does not usually exceed five and the number of organisms (or better, their biomass) diminishes rapidly. Although the species involved in a given food chain vary in space and time, similar sets of relationships recur in different habitats. For instance, the Arctic fox feeds on guillemot eggs in winter and seal carrion left by polar bears in summer; the spotted hyena has similar habits in Africa involving ostrich eggs and zebra killed by lions. Such professions were termed niches by C. Elton. *See* Biomass; Pyramid of numbers.

Food web. A neat classification by trophic levels often fails because a species may occupy more than one level during its life cycle or in response to changes in availability of food For instance, hover flies change from decomposers to herbivores,

chain.

Quantitative studies. In view of the complexity of food chains, attempts have been made to measure the size of the feeding links between species in order to select the most important. The only universal currency for such measurements is the caloric content of the food (*see* Ecological systems, energy in). A number of quantitative food-chain studies of this type permit comparison between different levels in a chain and also the chains in different communities. However, it must be stressed that certain species have an importance in the community which is out of proportion to their energy intake; for instance, the removal of pollinating insects may greatly affect the productivity of orchards and meadows; the excretion of external metabolites and antibiotics in minute amounts may



Food cycle among the animals on Bear Island, a barren spot in the Arctic zone, south of Spitsbergen (dotted lines represent probable food relations not yet proved). To read the diagram, start at marine animals and follow the arrows. (*From C. S. Elton, Animal Ecology, 1927*)

change the balance among microorganisms. *See* ECOLOGY. [A.MC.]

Bibliography: C. S. Elton, *Animal Ecology*, 1927; E. P. Odum, *Fundamentals of Ecology*, 2d ed., 1959.

Food engineering

The technical discipline involved in food manufacturing and refined foods processing. It encompasses the practical application of food science in the efficient industrial production, packaging, storing, and physical distribution of nutritious and convenient foods that are uniform in quality, palatable, and safe. Controlled biological, chemical, and physical processes and the planning, design, construction, and operation of food factories and processes are usually involved. *See* FOOD MANUFACTURING.

Food engineering is the food industry equivalent of chemical engineering. Food science in industry converts agricultural materials into products that are marketable because they meet a consumer need and can be profitably sold at reasonable prices by virtue of being economically produced, packaged, and distributed.

Food engineering is a vital link, therefore, between farms and food stores in the lifeline of modern civilization. Without it, food would be available only at farms, in forms produced by nature, and available only in season.

Many natural foods must be preserved in order to be stored and shipped long distances to population centers. To provide a varied, nutritious diet, many raw food materials must be refined, and others combined with various ingredients and processed to produce food in new forms, such as bread, cheese, ice cream, frankfurters, soft drinks, cake and dessert mixes, candy, many breakfast cereals, ketchup, and salad dressings.

Because food engineering is applied in food manufacturing and refined foods processing, it requires a knowledge of unit operations and processes such as cleaning, separating, mixing, forming, heat transfer, moisture removal, fermenting, curing, packaging, and materials handling. This requirement to a large extent differentiates food engineering from food science. Yet these operations involve applied food science, just as chemical engineering encompasses chemistry. That is why the food engineer must have a working knowledge of food chemistry, bacteriology, and industrial microbiology, as well as of physics, mathematics, and basic engineering disciplines.

Food engineering should be differentiated from food technology. In teaching technology, some schools include considerable engineering in the curriculum to train students in the practical application of food science. Graduates of these schools are in reality food engineers. Other schools emphasize food science, and qualify graduates primarily for research and quality control work. The food technologists from these schools actually are food scientists.

Illinois Institute of Technology gives a degree in food engineering. Its curriculum places particular emphasis on engineering principles and unit operations and processes. More than a score of well-known colleges and universities give degrees in food technology. Some of these, such as Massachusetts Institute of Technology, include much engineering in their courses.

Although courses are not standardized or accredited by any official body, the Committee on Education of the Institute of Food Technologists has recommended sound training in mathematics, physics, chemistry, and engineering principles.

Students interested in research select a curriculum which emphasizes food chemistry, bacteriology, and microbiology. Those interested in food manufacturing and processing choose a curriculum with less emphasis on food science and more on mathematics, physics, engineering principles, and unit operations and processes.

Food engineers are hired not only by commercial food manufacturing and processing firms, but also by food machinery and ingredient companies, makers of containers and packaging materials, government food control and research agencies, food laboratories, and schools teaching food engineering, technology, and science while conducting research on food and nutrition.

Food engineering has contributed to ever-accelerating progress in the food industry by advancing food processing and manufacture beyond the pot and kettle stage of batch operations or large-scale kitchen techniques. It has developed many automatically controlled, continuous processes which are more efficient, provide better quality control, and are more sanitary. It has developed food manufacturing and processing into a highly technical industry.

The food industry faces the challenge of practically taking the homemaker out of the kitchen by providing economical products that require almost no effort or time in preparation. Products which can be made ready to eat by being boiled or toasted in their packages go a long way toward attaining this goal.

Some outstanding achievements in food engineering include continuous bread-dough making and forming, manufacture of low-cost, high-quality prepared mixes, development of instant coffee and tea processes, dehydration of potatoes to produce the instant mashed product, production of precooked frozen convenience foods, continuous butter churning, conversion of juices to flake and powder form, pneumatic bulk handling of dry and liquid raw materials, and automatic control of processes.

Promising projects now under development are preservation of foods by nuclear or electronic radiation, heat processing by high-frequency radio waves, and freeze-drying of meat and seafood.

REFINED-FOOD PROCESSING

Refining of foods through processing is a large, important commercial business, vital to the food supply. Refining involves operations which convert raw materials into forms more suitable as food, with longer storage and shelf life.

Sugar, for example, is obtained by refining. Cane sugar is produced by squeezing the juice from the cane by the mechanical pressure of huge steel rolls; the juice is clarified and concentrated in big vacuum evaporators; crystals are formed in the concentrate, separated from the liquor by centrifuging, and then dried and sometimes pressed into cubes. The refining operation has merely extracted naturally occurring sugar from its source to make it usable in convenient form. *See* SUGAR.

Milk is clarified and pasteurized to make it safe to drink; it usually is vitaminized to improve its nutritive value. Soybeans must be refined to yield edible oil; corn to yield oil, starch, sugar, and syrup. Wheat is refined by milling to make flour. Similar processes apply to a long list of common foodstuffs. *See* MILK, SOYBEAN.

In addition to producing consumer items, refiners supply large quantities of their products in bulk to food manufacturers to be used as ingredients in manufactured foods.

Examples of refined foods are dried and frozen eggs, spices and flavorings, edible fats and oils, sugar, syrup and molasses, honey, starch, gums, gelatin, cocoa, coffee, tea, vinegar, milk, flour, rice, and oat cereal. *See* EGG PROCESSING; GELATIN; GUM; STARCH.

Food refining and manufacture together constitute not only a vital industry, but one of the largest ones in the United States. Food science, technology, and engineering have made refining operations highly technical and efficient, and are accelerating progress. A wide variety of continuous, automatically controlled unit operations and processes is employed to save labor, improve quality, and reduce costs

An example of significant advance is impact-in-air milling of wheat, in which the grain is disintegrated by high-speed revolving blades. The different fractions of the grain are then separated by air classification to yield flour with special baking properties. Another example is the tailoring of shortening to specific uses by modifying the molecular structure through chemical processing.

Many food refining plants are located near the area in which the raw material is produced, such as the soybean plants in Illinois, the cane sugar factories in Louisiana, and the flour mills in Minneapolis. Milk pasteurizing plants, however, are near their customers, because the milk is bott[

immediately after pasteurization. Food manufacturing plants producing highly perishable items, such as bread, usually are near their market. Those turning out long-life items, such as preserves, are more often near the raw material supply. Perishability of the raw material, as with fruits and vegetables for canning and freezing, often dictates that the food plant be close to where the food is grown. *See* COCOA POWDER AND CHOCOLATE; COFFEE; RICE; TEA; VINEGAR; *see also* BARLEY; BUCKWHEAT; OATS; PEANUT; RYE; WHEAT.

UNIT OPERATIONS AND PROCESSES

Unit operations in food factories are the mechanical manipulations and handling employed to change the physical form or composition of the food, to move it from one operation or process to another, and to package it.

Unit processes are the methods used to change the chemical or biological characteristics of the food, to preserve it, as in curing meat; to make it more palatable, as in aging cheese; or to develop special qualities, as in fermenting wort to produce beer by developing alcohol and carbon dioxide.

In the food industry, the two terms often are loosely employed to designate steps in the processing and manufacture of food.

A typical food factory performs a series of unit operations and processes in refining a raw material or in producing a fabricated food item. The combination of operations and processes is different for the various branches of the industry, but many of the individual operations and processes are employed in several branches of food treatment.

Washing, separating, mixing, scaling, handling, process control, sanitation, waste disposal, and packaging are examples of unit operations found in many types of food factories.

Application of heat, fermenting, curing, and aging exemplify unit processes common to various branches of the industry.

Since the widely used unit operations and processes are technical common denominators of the industry, knowledge of the engineering and scientific principles involved is essential to food engineers and technologists who go into the production phase of the industry.

Significant improvements in many unit operations and processes have taken place since 1945. Most notable is the trend to continuous instead of batch operations and processes, with improved efficiency, capacity, and controllability. Extensive application of automatic control has been part of the trend to continuous methods.

In heat processing, progress has been made through high-temperature, short-time pasteurizing and sterilizing. Not only is capacity increased, but there is less heat damage to flavor and vitamin content.

Notable advancement in handling has been achieved through extensive application of pneumatic conveying of dry materials and pumping of fluid and plastic materials.

Potentially important new processes in the development stage are (1) freeze-drying, or removal of moisture under high vacuum while the product is in a frozen state, (2) pasteurization and sterilization by irradiation with nuclear rays or electron beams, and (3) heat processing with high-frequency radio waves. *See* FOOD SCIENCE; FOOD TECHNOLOGY.

General operations. General operations are the processes employed in every type of food processing and manufacturing plant—and in many phases of plant operations—in contrast with specific processes. Handling of materials is an example. No matter what type of plant is involved, this operation is widely utilized. It is necessary to move raw materials, ingredients, and supplies from the receiving dock into plant storage, through the various process operations required to make the finished product, and through the packaging lines. Finally, the packaged product must be transported to the plant warehouse or directly to the shipping platform, and then moved into trucks or railway cars. For efficiency all of this is done mechanically.

Process control, sanitation, quality control, and waste disposal are other operations widely employed in every food plant. [F.K.L.]

Process control. Measuring instruments and integrated systems of devices are used to regulate automatically the chemical, physical, and biological changes which occur in food during manufacture. Application of food process control systems has a twofold objective, to preserve desired qualities in the product, and to achieve the lowest cost of operation.

Automatic control of food quality is maintained indirectly in most processes by measuring the important variables such as temperature, pressure, and level and flow of product and by holding these at desired values through regulating valves and allied devices. Subsequent laboratory tests of food quality then determine the effectiveness of control settings. In recent years the speedup of processes has necessitated a more direct method of control. In-process quality measurements have been developed for pH, color, viscosity, and moisture content and for special values such as iodine number in fats and sugar content in concentrated juices (Fig. 1). Usually these continuous measurements guide the process operator in setting his conventional controls. But as the dynamic characteristics of each process are learned, direct automatic control, bypassing the operator, is achieved.

More economical food processing is resulting from automatic systems which increase the flow of product through the process. Electromechanical devices are used to sort, count, and continuously weigh raw materials in process feed and blending systems. Similar equipment is used in the automatic high-speed packaging of foods. In many plants these automatic systems of feed and packaging are integrated with the process proper, with automatic control being exerted from a central instrument panel. [L.E.S.]

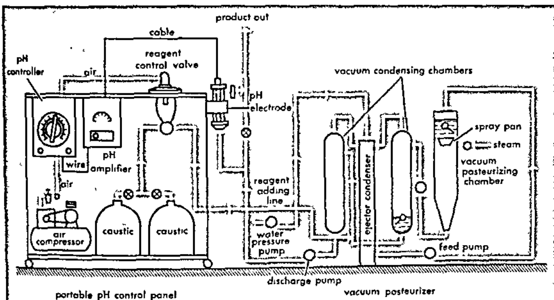


Fig 1. Robichaux continuous neutralization system. Acidity of pasteurized cream is determined continuously by pH electrode, then pH controller actuates reagent control valve, which dispatches required amount of caustic to high-vacuum chamber of pasteurizer. (Food Engineering, McGraw-Hill)

Sanitation. Sanitation in the food industry is the planned control of the production environment, equipment, and personnel to prevent or minimize spoilage, product contamination, and conditions offensive to the esthetic senses of the discriminating consumer, and to provide clean, healthful, and safe working conditions.

To be successful, competent management must plan and execute sanitation controls. The most important contribution of the sanitation program is the recognition and interpretation of potential hazards and their correction or control. The following general categories indicate some of the broad areas of sanitation concern in food industries.

Housekeeping implies orderliness and freedom from refuse in all areas.

Rodent elimination involves knowledge of rodent habits, recognition of problems, and permanent control through structural changes, removal of harborages and food supplies, and supplementary poisoning and trapping.

Insect elimination from food products and ingredients in the factory requires recognition of serious or incipient infestations, identification, and knowledge of habits and ecology. Control methods may involve changes in structure, equipment, or process, and the safe use of insecticidal chemicals.

Microorganisms, the type and significance of which vary with product and type of operation, must often be controlled by process and equipment change, cleaning, and sanitizing chemicals.

Construction and maintenance of buildings and equipment are of major importance in sanitation. New units can be planned to simplify sanitation maintenance, reduce costs, and eliminate the hazards of contamination and spoilage.

Cleaning of plant and equipment involves careful organization, training, work scheduling, and the use of the best available equipment, methods, and materials.

Employee facilities, such as rest rooms, locker rooms, drinking water, eating facilities, and working environment, must be well maintained for the comfort and safety of the workers if they are to remain happy and maintain production efficiency and product quality.

Laboratory tests, of importance to the sanitation

tems, as well as waste treatment and disposal, and lighting and ventilation, are often a part of sanitation.

Inspection techniques, tailored to the specific sanitation situation, must be learned, taught, and applied for efficient functioning and adjustment of the sanitation program. [E.S.D.]

Quality control. Quality control is the evaluation of raw materials, unit operations, unit processes, or finished products and comparison of the results with fixed standards. These standards may reflect the manufacturers' or the customers' viewpoints. They may be based on physical properties such as size and color, chemical properties like acidity, sensory attributes like odor and flavor, legal requirements like net weight, or on public health standards of microbial content. When possible, quality control depends on objective physical or chemical tests, but for foods these are usually supplemented by a panel of trained tasters. The tests are applied according to a statistical design, following an analysis of the specific problem (see EXPERIMENT). Frequency of sampling and repr

bility of results vary with the degree of quality control desired (*see* QUALITY CONTROL). Statistical analysis of data previously obtained defines the limits beyond which a product is rejected, an operation readjusted, or a process changed. In addition, good quality control can show trends that indicate changes should be made before any losses have actually occurred. Care should be taken that the quality control program does not become so detailed and costly that it reaches the point of diminishing return. Quality control is essential for the product uniformity that enables mass-produced foods to be advertised, distributed, and sold throughout the world. [R.T.M.]

Waste disposal. Both solid and liquid wastes are encountered in the food industry. The solid wastes are segregated and, in many cases, processed into salable by-products. Liquids may be discharged into streams or municipal sewers, or treated on the premises.

Legislation concerning pollution of surface waters has greatly reduced the quantity of liquids discharged into streams. Discharge into municipal sewer systems offers the best solution and is most commonly employed.

Fruit and vegetable canneries and freezing plants that have highly seasonal operations use large quantities of water that tend to overload municipal facilities. Many of these plants have found it necessary to install waste-treating systems. Most common methods employed involve screening, lagoons, septic tanks or cesspools, sedimentation, land irrigation, sand filtration, chemical treatment, and spray irrigation (Fig. 2). Screening is an absolute necessity unless wastes are discharged to lagoons where odors are no problem.

Milk processing, brewing, meat packing, and corn products refining plants each present special problems. However, good housekeeping practices, recovery of by-products, and reuse of waters can greatly reduce volume and concentration of the waste (Fig. 3). *See* SEWAGE TREATMENT. [A.V.C.]

Materials handling. Materials handling is the in-process and in-plant handling or conveyance of raw, semifinished, and finished materials to storage and point of shipment. Improved handling systems present the greatest potential means for cutting production costs, increasing production within present plant areas, and providing a smooth, continuous flow-through process to storage or shipment areas.

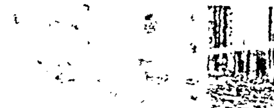


Fig. 2. Spray irrigation system. Screened effluent from vegetable processing plant is spread over forest area where it is quickly absorbed. Purified water then discharges to stream. (*Food Engineering, McGraw-Hill*)

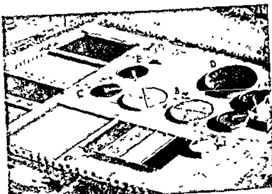


Fig. 3. Brewery waste disposal system. Brewery wastes have high BOD (biochemical oxygen demand) and suspended solids, so require extensive treatment. In this installation, liquid waste is pumped to primary clarifiers, A. Effluent goes through bio-filters, B, intermediate clarifier, C, secondary bio-filter, D, and clarifier, E before it is discharged. (*Food Engineering, McGraw-Hill*)

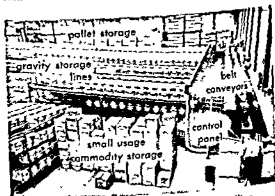


Fig. 4. Gallo wine electromechanical handling

Depending on a product's characteristics, it may be conveyed by semimechanical, mechanical, gravity, hydraulic, or pneumatic systems, or a combination of these.

In-process handling systems provide the conveying linkage between the unit operations of a process. With a few exceptions, such as monorail equipment used for meat carcass handling, most systems are designed for mass movement of materials. In all cases, every consideration is given to the product's characteristics, and systems are designed to handle materials with minimum damage, move them swiftly to minimum distances, and provide sanitary requirements.

In-plant handling systems include the conveyance of raw materials from receiving platform or storage to process, as well as the handling and storage of finished goods (Fig. 4). The mechanized means selected for these purposes also are based on consideration of the characteristics of the raw material and that of the finished product's container and its shipping package. *See* CONVEYING MACHINES.

Flow rate, time, quantity, kind and direction of flow may be automatically controlled by instruments, electromechanical or electronic devices and systems. [C.B.H.]

Raw materials preparation. The processes discussed in this section are those operations necessary to make natural food materials from the farm ready for marketing or for further processing.

Citrus fruit, for example, may be washed, sorted for size, inspected, waxed, and even colored before it is shipped to fresh-fruit markets. Much fresh produce is cleaned, inspected, sorted, trimmed, and prepackaged for retail sale. This assures better quality, reduces waste, is sanitary, enhances sales appeal, and saves labor in food stores.

Citrus fruit also may be cited as an example of a raw material that is prepared for processing. The fruit is inspected, cleaned by a washing operation, and sorted into size classifications before the juice is extracted, concentrated, and frozen.

Unit operations employed in raw materials preparation include cleaning, separating, draining, trimming, peeling, dehusking, hulling, shelling, pitting, coring, stemming, silking and cutting, centrifuging, dressing, filtering, solvent extraction, and rendering.

These operations do not change the basic characteristics of the food material. This is in contrast with raw materials conversion, which does change the nature of the material. [F.K.L.]

Cleaning. The cleaning operation consists of the removal of soil, dust, spray residues, insects, superficial rots, and other contaminating substances. This is accomplished by tank soaking, pressure sprays, brush and shaker washers, and detergents. Even acids and alkalies have been used for removing spray residues from fruits and vegetables. The United States Food and Drug Administration considers that commercial cleansing methods should be fully as effective as good home kitchen practice. Both blowers and vacuum suction are used to remove debris, dust, and insects from dry or semidry food materials like cereals. Magnets placed in the processing lines are commonly used to remove metallic objects from such dry materials as beans, rice, and cereals.

Separating. Separating is a unit operation used to classify products as to color, size, weight, shape, and texture. There are a number of separators, one of which is the gravity type which depends upon the flow of dry materials on an inclined plane. Size separations are easily made by passing peas, for example, down an inclined plane fitted with openings of several sizes. Another is the quality-gravity pea separator which separates peas on the basis of weight; the more mature peas, being heaviest, sink. Electronic devices placed on a belt conveyor will separate and discard dark-colored particles, as in breakfast foods, or discolored grains in rice or cereals. The so-called squirrel cage washer and separator consists of an inclined rotating and sometimes vibrating screen which separates the peas into four or five sizes.

Roller sorters are accurate and fast and do not injure soft fruits such as pears and apples. Roller conveyors with a fixed space between the rolls will remove debris such as stones, soil clumps, and leaves. Weight sorters have been perfected to the point where they are reasonably fast and accurate and operate without injury to such soft fruits as tomatoes, pears, and apples. By far the most common separators for grains are the several vibrating screen devices, often fitted with gravity flow and air blasts. Screens in series and with different sized perforations are commonly used. So-called fanning mills for small seeds fall into this class.

Centrifuges are used to separate fat from milk or other emulsions. They are also used to remove sediment or precipitates from liquids such as fruit juices. Cyclone air separators for fine particles like flour, meals, small seeds, and powdered foods are used in many designs. In discussing the separating operation it must be kept in mind that several types of separation are normally used in the processing of a single food.

Flotation is a physical principle often used in separating products of different specific gravity. Examples are the quality-gravity separat[illegible]

separated from the pits by flotation in running water. [illegible]

passes through brine flumes fitted with baffles and a good separation is obtained. In machine-extracted crab meat a good separation of chitinous materials and shell from the meat is obtained by brine flotation.

Draining. In the preparation of fruit, vegetables, meat, and seafood for canning and freezing, careful and consistent draining is required so that excess water is not transferred to the can or package. Food laws severely restrict the quantity of water so transferred. Thus peach, apricot, and pear halves are always placed cup down on belts to drain before filling. Cherries, berries, and other foods which absorb water easily must be well drained on vibrating screens or belts to remove excess moisture. Only by using fruit of a constant moisture content can the siruping operation be accurate, that is, the maintenance of desired sugar concentration, or Brix, in the consumer product. If oysters and clams are not drained carefully, it is impossible to secure the required drained weights in the can. On the other hand, in a product like tomatoes, draining can be excessive and much of the juice which is an integral part of the tomato may be lost. Care must be taken in the blanching operation to avoid undue loss of moisture from such vegetables as spinach, broccoli, and snap beans.

Trimming. Trimming is a simple operation usually performed in one of two stages during the preparation of vegetables or fruit for processing. Crude or preliminary trimming is often done before the product is washed. It is a part of the sorting operation and consists in discarding culls, and mis-

shapen, discolored, semidecayed specimens. The tops of vegetables, as well as discolored and slightly decayed portions, are removed by hand trimming. More careful trimming is done on a conveyor belt after the fruits or vegetables have been carefully sorted and cleansed. Here, surface blemishes, immature or bruised portions, and pieces of skin are removed. Such trimmed produce is often consigned to a lower grade, but it is still sound food and need not be discarded in spite of its poorer quality and sale value.

Peeling. Peeling is done both by hand and mechanically. Hand peeling has been largely replaced by steam, abrasion, gas flame, or lye treatment to remove the peel Some fruits, such as plums, figs, cherries, berries, and apricots, are not peeled at all. Although there are no standardized methods of removing the peel from such dissimilar foods as pineapples, grapefruit, apples, pears, peppers, and sweet potatoes, equipment is available for handling each product. Hot lye (sodium hydroxide) solution is used for peeling peaches, pears, sweet potatoes, carrots, and peppers. A low concentration of 1.5–3% is used for fruit while from 5–15% is used for vegetables. The albedo or rag of peeled grapefruit is usually removed by a spray or a dip in lye solution, the lye being promptly removed by a water spray or dip. There is little or no residue of the lye on products that have been treated with it. Since lye tends to discolor peaches and some other foods, acid washes are sometimes used to preserve the natural color. Steam peeling is used generally for white potatoes and some other vegetables since it is a less costly method for large quantities. It also gives the highest percentage recovery of edible portion. Hand trimming is usually necessary to remove minor defects. Abrasive-type peelers are used today mainly in restaurants and institutions since they are wasteful and leave an undesirable rough texture. Roasting by gas flame is an unusual method which is sometimes used for peeling sweet [illegible]

[illegible] all the larger canning and freezing plants use efficient and speedy mechanical huskers, silkers, and cutters. Huskers consist of a pair of rapidly revolving rubber or milled-steel rolls which catch the husks much as a clothes wringer does wet clothes. The rolls can spread apart sufficiently to allow the husks to pass between them, but powerful springs return them at once to their normal position. Single machines operate at the rate of 60–70 ears a minute, and the twin-type machines double this capacity. Most corn factories use a desilker which consists of a series of revolving rolls and brushes which remove the silk before the corn is cut from the cob. Final desilking takes place in the flotation separator following the cutting operation. There are several types of corn cutters. The husked and trimmed ears are fed through curved knives that accomodate themselves by springs to the size of the ear. For cream-style corn, scrapers remove the remainder of the kernel still clinging to the cob after the first cutting process. Some packers pulp or Creamogenize a portion of the corn and return it to the canned product to improve its consistency. The present trend is a preference for whole-kernel canned and frozen sweet corn. The whole kernels are removed from the cob by specially designed cutting knives. The kernels are carefully inspected on a conveyor belt and the silk and debris are removed by a flotation separator. A dewatering vibrating screen removes excess moisture before the kernels pass to the filling machine.

Shelling. Shelling is an operation limited to raw foods such as dry corn, beans, dry peas, peanuts, pecans, and walnuts. Corn is readily shelled in one of several types of contact shellers. Often the process includes air-blast cleaning to remove adhering chaff and [illegible] should [illegible] contain no [illegible] hominy is [illegible] to loosen the hull, leaching out the lye, and cooking the greatly enlarged kernels until tender. *Hulled* corn was a very important colonial food but is little used at present.

Fresh peas, chick peas, and browneye peas are shucked by threshers or viners located either in the fields or at the processing plant. The viners simply beat out the peas from the pods. Both peanuts and walnuts are shelled by passing the nuts between carefully adjusted rollers which crush the shell but do not break the meats. Popcorn is normally dried to a moisture content of 13–14% before shelling This is the moisture content which gives maximum popability.

Stemming and pitting. Stemming and pitting are necessary for some fruits, especially cherries While cherries are usually stemmed by hand, machines are available which save time. Blueberries harvested with scoops contain stems which are difficult to remove. The use of vibrating screens helps to remove many stems but no method has been found to be fully effective. Blanching blueberries for 10–30 seconds helps to remove the stems and also tenderizes the skins for freezing. *Gooseberries* may be passed through a machine resembling a bean snipper to remove stems and floral cups.

Sweet cherries are not usually pitted. Sour or pie cherries, and to some extent dates, olives, and raisins, are pitted by forcing a 3-pronged plunger through firmly held fruit. Raisins are steamed before seeding. Apricots and peaches are normally hand-pitted but pitting machines are available which greatly speed up the operation.

Coring. Coring is limited to such fruits as apples and pears. Apples are usually run through a combination peeler-corer and seed-cell remover, an ingenious combination of three machines. Manual trimming is necessary after machine peeling and coring no matter whether the apples are to be canned, frozen, or used in applesauce. Seed cells constitute a serious defect and must be removed Pears are cored either by hand or by one of several

types of mechanical corers. A special contour-bladed knife is used in hand coring. Because of the peculiar shape of the pear, considerable hand trimming is necessary. [C.R.F.]

Filtering. This technique separates liquids and solids by means of a pressure differential which may be created either by vacuum on the underside of filtering media, or by external force applied to the filter cake. The technique is largely used in breweries, wineries, fruit-juice plants, and vegetable-oil and sugar refineries.

Pressure filters employ papers or cloth supported on horizontal or vertical leaves, or fine metal screens with 2500 or more openings per square inch. These are usually precoated with nonfibrous materials such as diatomaceous earth (Fig. 5).

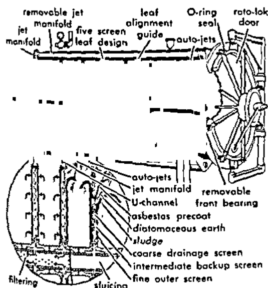


Fig. 5. Cutaway drawing of vertical leaf pressure filter. Slurry enters side of shell and is evenly distributed to the leaves where solids are deposited. Filtrate moves to center shaft and is discharged. For cleaning, water is sprayed on slowly rotating leaves from header in top of shell. (*Food Engineering, McGraw-Hill*)

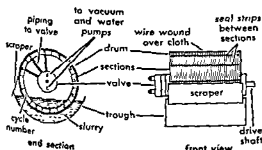


Fig 6. Drawing of vacuum filter. Vacuum filter comprises compartmented drum rotating in trough. Slurry is fed to trough and picked up by drum. Liquid portion is drawn through filtering media and discharged at center. Solids are deposited on surface of drum and removed by scraper.



Fig 7. Beef carcasses are shrouded with a muslin cloth as they leave dressing floor on way to cooler. The shroud helps to smooth and whiten the fat. (*Armour and Co.*)

Most vacuum filters comprise compartmented drums covered with woven wire cloth, perforated metal, cotton, wool, or synthetic materials (Fig. 6). The drum is rotated in a tank containing slurry to be filtered. Vacuum, applied to the inside of the drum, picks up slurry and deposits solids on filter media. The filtrate discharges through a valve in the axis. As the drum rotates, the cake may be washed and partially dried by vacuum before it is removed by a scraper.

For slimy materials, drums are precoated before operation and, as they rotate, a moving knife continuously shaves off deposited solids plus a small amount of the filtering media. [A.V.G.]

Dressing. Slaughter and dressing are generally a continuous process, except where animals are frozen prior to dressing. Dressing includes all operations required to prepare the hot carcass for chilling (to remove body heat) and subsequent breakdown.

The degree of dressing depends upon the type of animal, market requirements, local preference, and custom. Dressing on the rail is the most modern procedure employed today.

Shipper beef is eviscerated, hide removed, head and feet dropped, split into sides, and shrouded for chilling (Fig. 7).

Dressing of sheep is quite similar; calves, however, are dressed and chilled with the hide on or off.

Hogs are scalded and singed to remove hair prior to evisceration and splitting (Fig. 8).

Fish may be completely dressed, scalded, and



Fig. 8. Hog-dressing line in Chicago meat packing plant. (*Armour and Co*)

Expressing. Expressing is a mechanical process for separating the liquid and solid components of a substance by the use of pressure. The process is widely used in separating the oil from such oil-bearing material as oilseeds, nuts, fruits, and fatty animal tissue and to a lesser extent in the separation of juice from such fruits as citrus fruit, grapes, pineapples, apples, and tomatoes.

A great variety of [illegible]

[illegible] cally, or centrifugally.

Materials to be expressed are normally pretreated, ground, and heat-treated in the production of oil. Juice fruits are ground or crushed.

Solvent extraction. This is a process for separating one component of a substance by use of a solvent in which the component is soluble. *See* SOLVENT EXTRACTION.

The process is widely used in separating oil from oil-bearing material, principally and most practically when the oil content is low, as in soybeans. Solvent extraction is also used commercially in processing cottonseed, flaxseed, corn germ, wheat germ, animal fat, castor beans, soybeans, and cocoa, and for the extraction of essential oils (Fig. 9).

After pretreatment (expressing, grinding, crushing, roasting, or flaking), materials to be extracted are brought in contact with a solvent in either batch or continuous processing equipment. Solvents normally used are low-boiling hydrocarbons or chlorinated hydrocarbons. The solvent extract or miscella is distilled to remove the solvent. The extracted cake is also desolventized.

Rendering. Rendering is a process employing heat for separating oil and fat from animal tissue. Heat is applied principally to coagulate cell proteins, permitting release of the oil. Rendering is universally used in the production of lard and tallow from meat fats.

Two rendering processes are used. Dry rendering is carried out either in open agitated kettles or in closed vessels under vacuum. Wet rendering is con-

1 bean storage
2 bean cleaner
3 magnetic separator
4 surge bin
5 scale
6 cracking roll
7 bean heater
8 flaking rolls
9 elevator to extractor
10 extractor filling hopper
11 extractor
12 extractor baskets
13 spent flake conveyor
14 desolventizer toaster
15 flake cooler
16 solvent pump
17 half miscella pump
18 full miscella pump
19 miscella filter
20 heat exchanger
21 pre-evaporator
22 entrainment separator
23 condenser
24 vacuum stripping column
25 finished oil pump
26 solvent work and water separation tank
27 waste water evaporator
28 cyclone
29 rotary valve
30 meal screen
31 meal grinder
32 solvent surge tank
33 half miscella surge tank
34 hydraulically operated valve

Fig. 9. Flow chart of soybean preparation, solvent extraction, solvent recovery, and meal grinding. (*French Oil Mill Machinery Co.*)

ducted in open kettles at low temperatures with added water, or at high temperatures under steam pressure. The type of oil or fat product desired usually determines the type of process used. [C.K.WI.]

Centrifuging. Centrifuging is a technique for separating solids from liquids, or liquids from liquids by application of centrifugal force. It is widely used in the food industry for clarifying fruit and vegetable juices, milk, and oils; separating fat from milk and starch from gluten; dewatering crystals; reclaiming meals and oils from meat and fish wastes; and for removing sludge from hot wort.

Two basic type units are employed in the operation, settling machines and filtering centrifuges. In settling machines the mixture of liquids, or of liquids and solids, is stratified by centrifugal force. The separated components are either drawn off continuously, or one settles on the inside of the bowl and is removed intermittently by cleaning (Fig. 10).

In filtering centrifuges, solids are collected on the surface of a perforated basket, while the liquid passes through the accumulated mass and is caught in an annular casing which diverts it to a discharge point (Fig. 11). Provision is made for purifying the separated crystals by washing. Solids may be discharged from the basket intermittently or continuously. [A.V.C.]

Raw materials conversion. This is the processing of natural foods from the farm to preserve them,

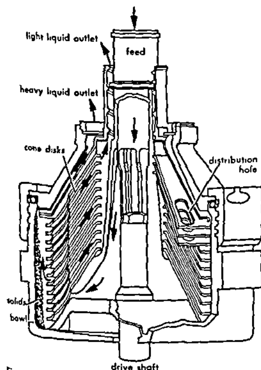


Fig. 10. Settling centrifuge. (Food Engineering, McGraw-Hill)

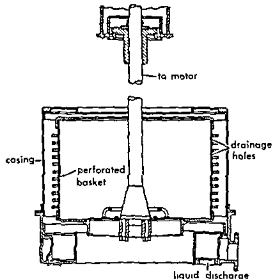


Fig. 11. Filtering centrifuge. Perforated baskets are employed. Crystals are deposited on inside, liquid moves through openings, and is discharged from casing. (Food Engineering, McGraw-Hill)

change them into a more useful form, or combine them with other materials and ingredients to produce manufactured foods. Examples are milling wheat into flour, fermenting cabbage to make sauerkraut, smoking meat, and making bread.

Additional examples of conversion operations include size reduction (such as cutting and grinding), *mixing* (to combine various materials and ingredients), forming into a particular shape and size (such as fishsticks), heating (as in cooking preserves), freezing, and dehydrating. [F.K.L.]

Size reduction The mechanical size reduction of materials to facilitate processing by the production of slurries, pastes, and particles with distinct size or shape characteristics (for example, by grinding, cutting, comminuting, disintegrating, pulping, pureeing). The mills used are of sanitary design, with product surfaces and even entire units fabricated of stainless steel. Some have jacketed feed-throats and milling chambers for chilling or heating heat-sensitive or low-melting-point materials. The air-stream may be tempered to regulate final product temperatures or moisture levels; facilities for milling in an inert atmosphere may be separate or in conjunction with these special effects. Some processes utilize a prebreaker, preceding the mill, to reduce materials to acceptable size.

Grinding is done in modern attrition mills (plate, stone, stone-composition, or roller) which are mechanical descendents of old stone mills, but are now so refined that their end products, capacities, and efficiencies are highly improved. One such mill uses rotating (5400–7400 rpm) carborundum plates, grinds materials to smooth (0.0025- to 0.005-in. particle size) texture. A *roll-type coffee mill*, adaptable to multiple-unit sequential grin -

ing, employs micrometric roll adjustment in increments of 0.002 in. New grain milling principles include fluid-energy, turbo milling, and impact milling. Employing the principle of fluid-energy attrition, material is injected into a cylindrical reduction chamber of irregular, oval shape. It is entrained in a pressurized stream of fluid, air, or gas moving at supersonic velocities. Violent jet action shatters the grains by impact and abrasion. Centrifugal force shifts large particles to the outer periphery; smaller ones drift to the inside of the cylinder and, as they are reduced to submicron size, exit in the spent stream. An inclosed vertical rotor with radial blades segmented to its shaft provides an air-vortex principle for ultrafine grinding. The rotor turns at 1500–2500 rpm, creating air vortices between blades of 30,000-fpm peripheral speed. Particles of 1-micron size result from impact, in air, against walls; and from interparticle abrasion and attrition in air vortices. In impact milling of whole-grain flour, grain is batched into a horizontal, clockwise revolving drum (140 rpm); centrifugal force holds it against the inner periphery until it strikes a diversion bar. This hurls it into a gang of saws revolving counterclockwise (2400 rpm), which split the grain. The number of passes determines particle size.

Cutting is another method for reducing materials to desired configurations by halving, slicing, cubing, chopping, or comminuting. Halving, for example, of cling peaches is done by high-speed halving-pitting machines. One of these automatically positions peaches so halving knives divide them at the fruit's suture, then by torque action twists the halves apart, simultaneously releasing the pit from the flesh. Finer honing and induction hardening of knife edges allow high-speed, simultaneous, transverse-longitudinal cutting, producing cube shapes as small as ¾ in. An example of reducing material to pulp by cutting is the silent cutter in which a rotating bowl passes sausage meat under a rotating circular knife.

Comminuting-disintegrating machines have in common the mechanical principle by which they achieve size reduction. Fixed swing hammers or knives are attached to a rotor which revolves at high speed within a chamber. Some units operate on horizontal axes, others are vertical or set at decided angles (Fig. 12). Fine particles are separated and desired particle sizes are discharged through classifying screens, mechanical or air-separation devices. Coarse particles are recycled. Crystalline, granular, soft, or fibrous materials are reduced to the following sizes: coarse (20–80 mesh); fine (80–300 mesh); superfine (200–400 mesh); and ultrafine, down to 3–5 microns.

Pulping mechanically reduces materials such as meat, fruits, or vegetables to a moist, undissolved mass; and pureeing reduces materials of high moisture content such as fruits or vegetables to a pumpable slurry. These results may also be achieved by boiling. Often a unit process is composed of one or more of the preceding size reduction unit operations, for example, size reduction by special machine and then sieving the undesirable pulp from puree, or liquid from pulp. A familiar process is that used for tomato products in which a combined chopper-pump, or a comminuter, reduces whole tomatoes to make the hot-break process possible. [C R H]

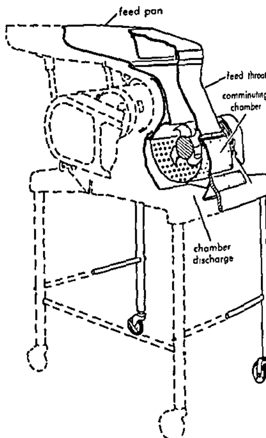


Fig. 12. Comminuting machine. (W. J. Fitzpatrick Co.)

Food mixing. Food mixing is a key unit operation involving instrumented blending of dry and liquid ingredients with or without heat (cooking) and on a batch or continuous basis.

An automatic system of making bread dough, supplanting operations of three bakery departments, has an hourly capacity of 4000–6000 loaves (Fig 13). The initial ingredients are agitated in a tank in which yeast activity is simulated. Flour is added and the liquid sponge or mix is pumped to agitator-equipped holding tanks. The sponge is then pumped to a continuous horizontal mixing trough into which liquid sugar is metered. It is finally pumped to a dough-development system, where dough pieces are extruded into pans that are conveyed to final proofer and oven.

Another example is the 2-step, continuous blending of mayonnaise and cake batters. Solids and liquids are metered into a bowl-type preblender with twin-disk impeller. The blend proceeds to a postblender, which is equipped with rotating and stationary toothed disks and a compressed gas inlet. In producing mayonnaise, the final mixing with vinegar is done in nitrogen atmosphere. By providing

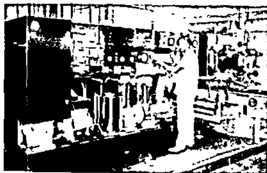


Fig. 13. Continuous bread-dough system. (American Machine and Foundry Co.)

starch-vinegar cooking-cooling facilities, the system also produces salad dressings.

Ultrasonics is offering rapid mixing, emulsifying and homogenizing of mayonnaise, dressings, cheese spreads, margarine, soups, sauces, ice cream, and peanut butter, with improved smoothness and stability.

Modern colloid mills now mix, mill, and homogenize, in single pass, to offer stable emulsions and uniform size reduction of such items as dressings, meats, fish pastes, cheese, and chocolate.

Dry or semidry mixes, like chocolate, powdered milk, and bakery premixes, are uniformly blended by a high-speed paddle-type unit that breaks up and disperses agglomerates. The unit can handle 3000–10,000 lb per hour.

Machines can automatically proportion ingredients into a dough mixer that extrudes and cuts macaroni, spaghetti, and noodles to lengths at a capacity of 1000 lb per hour.

A new conical mixer requires little power and employs a precessional motion of the rotating screw for rapid blending of dry materials and wetting or dissolving of solids.

A heated, rotating double-cone unit blends and dries products under vacuum. Another high-capacity rotary dry blender (4-way action unit) assures high mix uniformity.

High-volume blenders now speed the uniform mixing of dry materials and the dispersion of liquids, regardless of viscosity. Rotating twin-shell units (50 ft^3) have been modified for a liquid feed operation like blending melted shortening into dry mixes, and spices or flavoring emulsions into granular bases in 10–15 min.

An advanced starch-gum jelly process employs mixer-heat-exchanger assembly for continuous cooking and cooling of starch-sugar slurry at the rate of 2500 lb per hour. Before the cooked product goes into the cooling section, the flavor-color solution is introduced.

A simplified aerating process for making foamy marshmallow and nougats consists of (1) mixing air and liquid by jetting, (2) further mixing by mechanical impingement at high velocity against a fixed target from which the product is deflected with turbulence, and (3) straining and diffusing through a porous medium.

In fluid mixing, propeller- or turbine-type units do an efficient job with most liquids when impeller size, speed, and mounting are properly specified.

Food forming. Food forming plays a key role in mechanized shaping of products to precise shape and weight. A dozen innovations follow:

1. Because of continuously extruding dough in the automatic production of white bread, a machine with only two operators can extrude 70 loaves (1 lb each) per min (Fig. 14).

2. In those food industries requiring a product in individual pieces, a precision rotary unit can vacuum-form 2500 lb of product per hour. This permits continuous operation into a fryer, enrober, or freezer.

3. Mint creams are deposited onto a continuous belt and proceed directly to cooling and chocolate coating units.

4 Modern equipment hopper-feeds confections through sheeting rolls onto belts traveling through a chilling tunnel and then slits the sheet of multilayer confection into continuous strips that are cut into bars for chocolate coating. Cookie batter is similarly sheeted through rolls, with the sheet cut into shape. The batter may be hopper-fed through forming dies.

5. An air-operated unit can shape 400 link-type skinless sausages per hour. The machine also produces other meat emulsions in desired forms.

6 There is a machine that forms a wide range of products in club-shaped packages by shaping and sealing roll-fed plastic film into a tube into which the product is metered; package ends are sealed with metal clips.

7. Mold-type chocolates such as tablets, solid bars, and filled confections may be produced at the rate of 800 per min.

8. New equipment cuts fruits, vegetables, and meats into various forms such as strip, slice, dice, quarter, crinkle-cut or French-fry.

9. An automatic molder-twister-panner twists and joins dough pieces at a rate of 90 per min.

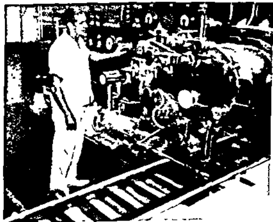


Fig. 14. Forming bread loaves. (Am and Foundry Co.)

10. A machine extrudes 1000 lb of dough per hour to form macaroni, spaghetti, and noodles.

11. There is a process for dry or press coating tablets as a substitute for costly pancoating operations.

12. Meat emulsions can be stuffed into casings in making frankfurters and other sausage products. Another more continuous sausage process consists of grinding meat, mixing in spices, and then pumping to a comminutor, which feeds to receiving tanks and stuffing nozzles. [J.V.Z.]

Heat treatment. Heat treatment of foods is an important part of the conditioning of foods for preservation between the time of harvest and the time of consumption.

Heat treatment follows many different patterns. It may involve application of heat indirectly, as in an evaporator or in a tubular heat exchanger, or it may be accomplished through direct contact of the heating medium with the food product. For a given purpose, either direct or indirect heating may be practiced, depending upon circumstances. For example, food sterilization usually is accomplished by heating the food in sealed containers of metal or glass; under certain circumstances, however, food is either sterilized or pasteurized while flowing through a tubular, plate, or other design of heat exchanger; or by batchwise indirect heating, as in a steam-jacketed vessel. In still other circumstances, food is sterilized by putting saturated steam into direct contact with the food, either in batch procedure or by moving the food continuously through a device in which either steam is injected into the food or the food passes through an atmosphere of steam.

Batchwise heat application is that in which the application of heat is sporadic or discontinuous. The food is treated in separate portions and each portion receives the required amount of heat to accomplish the intended purpose, after which the output of heat is stopped. Continuous processing, on the contrary, is that in which the supply of heat is continuous in a certain region or location and the food is moved continuously into and out of the influence of the heating medium within the given region. During this cycle of movement, each particle of food receives the amount of heat energy required to produce the necessary effect upon the food.

Commercial processing techniques are changing gradually from batch to continuous types. Continuous processes are more economical in the use of heat energy, labor, and time than are batchwise processes but cost of equipment for continuous processing generally is greater. The continuous food sterilizing system most widely used in the United States operates on only 50% as much steam and from 15–40% as much labor per unit of food processed as conventional batchwise systems (Fig. 15).

In Europe, two continuous systems, known respectively as hydrostatic sterilization and high velocity air sterilization, are prominent because their equipment cost is lower than that of other continuous systems for processing sealed containers. The hydrostatic system uses approximately 25% less steam and water per unit of food processed than the conventional batchwise system of processing (Fig. 16).

In systems in which low-acid foods are sterilized in sealed containers, temperatures varying from 121.1–135°C are commonly used. Pasteurization of milk is accomplished at 73.9°C. In continuous equipment for sterilizing food before it is packaged, food may be heated to 154.4°C (Fig. 17).

The effectiveness of certain bactericidal agents, such as irradiation and antibiotics, is being studied for use in combination with heat, with the objective of accomplishing either sterilization or pasteurization by heat treatment of reduced intensity.

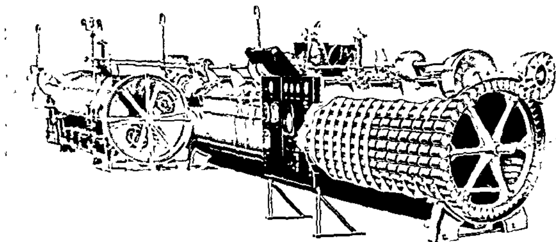


Fig. 15. Cutaway view of continuous pressure sterilizer and can-entrance valve of reel and spiral type. The smaller tank is a continuous pressure cooler. (Food Machinery and Chemical Corp.)

inlet water seal
outlet water seal
air cooling section
pre-heating section
bringing up leg
steam compartment temp 116.7° - 122.2° C
No. 2 feed and discharge mechanism No. 1
spray cooling leg
water seal temp 112.7° - 118.3° C
bringing down leg
floor level
cooling pit temp. 10° - 21.1° C

Fig. 16. Schematic diagram of hydrostatic system of sterilizing food in sealed containers. *(Mitchell Engineering, Ltd., London)*

Examples of temperatures used commerically in continuous bulk food heating for purposes other than either sterilization or pasteurization are the following: (1) in direct heating processes employing air as the heating medium—for flash drying of starch, 260°C; for baking of bread, 135–148.9°C; and for dehydration of foods, 65.6–93.3°C; or, in vacuum, lower temperature; (2) in direct heating processes employing fat as the heating medium—for deep-fat cooking of potato chips (Fig. 18), noodles, or doughnuts, 160–182.2°C; (3) in indirect heating processes employing steam—temperatures used for gelatinization of starch and for sugar inversion in jelly confections, 140.5°C; for evaporation to produce concentrated foods, 15.6–87.8°C; and for the rendering of lard in jacketed kettles 110–115.6°C.

The employment of direct heating for purposes other than sterilization or pasteurization includes also the steeping of rice in hot water under pressure as a step in making converted rice, the drying of food in high-velocity, high-temperature air as in making instant (precooked) rice or decaffeinated coffee beans, and the steam rendering of lard under pressure.

An additional example of an indirect heating process is the heating of cream to the desired temperature for the separating step in the process of butter making.

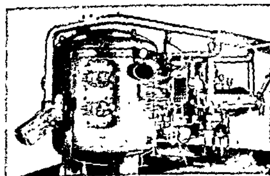


Fig. 17. Machine which sterilizes and cools food in bulk, then packages it in cans *(American Can Co.)*

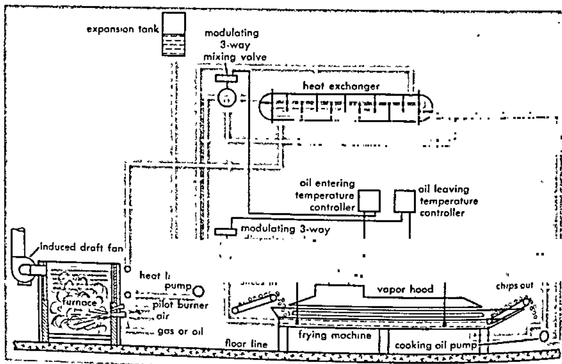


Fig. 18. Schematic layout of Smitherm process for deep-fat cooking, installation for potato chips. (H. W. Lay and Co., Inc.)

Cooling of foods. In processing this is a step in preparing food for commercial distribution to consumers. This phase of processing involves two distinct types of technique, applied respectively to two distinct categories of food.

The first consists of meats, dairy products, fruits, and vegetables which are marketed in the fresh state. To avoid serious loss of these articles through decay, they are substantially reduced in temperature, either before their transportation begins or during the first part of their transport period, and are held at a temperature between 0 and 4.4°C [illegible]

[illegible] refrigeration immediately after production, and milk products are similarly cooled as the last step in the manufacturing process. Packing fruits and vegetables with ice when loading for transport is still widely practiced in some localities.

Two systems of precooling, or cooling prior to shipping, which are generally regarded as embodying improvements over the old ice packing method for fruit and vegetables, are called hydrocooling and vacuum cooling. Hydrocooling consists of the continuous cooling of water, spraying it onto the product held in wholesale-size containers, and recirculating the water. The water generally has salt in solution to lower its freezing point and may also contain a fungicide or a bactericide, in which case the process is usually called stericooling (Fig. 19). Vacuum cooling performs its function by causing water to evaporate from the surfaces of the vegetable or fruit by producing a vacuum of 29.7 in. or more around the product. This system is especially good for products such as lettuce and spinach which have large surface in relation to volume.

During transport the temperature of cooled fresh meat, dairy products, vegetables, and fruit is maintained either by mechanical refrigeration or by bunker ice in the car or van. In cooling and in storing fresh produce, sufficient cooling capacity must be used to lower the temperature, counteract heat access by radiation, and counteract the vital heat (heat of respiration) of the product. To illustrate, a carload of tomatoes of specific heat 0.95 and weighing 20,800 lb must lose 19,760 British



Fig. 19. Stericooler in operation on peaches. (Food Machinery and Chemical Corp.)



Fig. 20. Vacuum cooler of railway car size. (Gay Engineering Corp.)

thermal units (Btu) per degree of cooling. Under average conditions, heat penetrating from outside the car amounts to approximately 3000 Btu per degree temperature differential, and vital heat per carload of ripe tomatoes, at 26.7, 21.1, 15.6, or 10°C, is approximately 103,000, 73,000, 54,000, or 34,000 Btu per day, respectively (Fig. 20).

Approximately 14% more heat is evolved from mature green tomatoes than from ripe tomatoes. Most fruits evolve heat at a higher rate than do tomatoes, while vegetables as a class evolve heat more rapidly than fruits. The rate of heat evolution by either vegetables or fruit varies roughly as the ratio of area of exposed surface to weight of material.

The second category of foods includes those that are heat treated in their processing and are commercially distributed in ways other than through restaurants. Only a minute percentage of these are sold to consumers in hot state. Therefore, most foods in commercial channels are cooled. Foods that are sterilized by heat are cooled after sterilization either by cooling the sealed containers with cold water or by passing the bulk food through water-cooled heat exchangers, depending upon whether the sterilization is accomplished by conventional means or by a system in which the food is sterilized before it is packaged. Foods pasteurized by heat are similarly cooled—either in the closed container or by heat exchanger. Cooling in closed containers may be by continuous, automatic operation, but more often it is by batchwise operation. Cooling of bulk material is predominantly by continuous operation.

Foods that are heat treated for purposes other than either sterilization or pasteurization and that receive cooling by a continuous operation as a step in their processing include starch jelly confections, cooled from 140.5 to 97.8°C by spreading on the surface of a cooled cylinder; chocolate which, during the pulverizing operation, is subjected to circulating air at 10°C to keep its temperature below 46.1°C; converted rice which, after a pressure steam treatment, is cooled in circulating air; citrus concentrate which is prepared by freezing out the water; butter which is chilled to 4.4°C in ammonia-cooled cylinders; and foods of many varieties in cartons, including ice cream, which are frozen in numerous types of continuous systems.

[C.O.B.]

Freezing. This is the use of subzero temperature for the freezing-preservation of foods, which thereafter are stored in low-temperature environment until consumed. Freezing is achieved by automatic plate freezers, continuous air-blast units, direct-immersion freezers, continuous freezers for liquid and semiliquids, and batch units; examples are the double-contact pressure plate, tray cart in air-blast tunnel, and air-blast room freezers.

Moisture removal. The preservation of foods by sun-drying, the oldest known food process, is still practiced. However, modern principles of evaporation as applied to semisolid and liquid food materials, new thermodynamic techniques, and new unit operations are utilized to produce a number of new liquid, solid, or powdered foods.

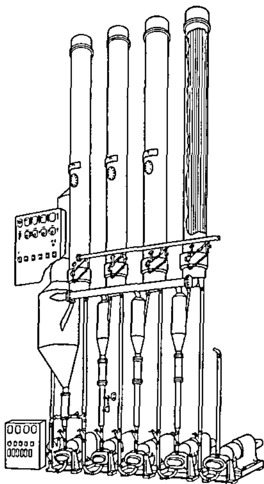


Fig. 21. Thermovac, 4-stage, single-pass, highly automated evaporator. Temperatures and velocities are minutely, accurately, and automatically controlled in all four stages at upper and lower control panels.

Evaporation or boiling down, the classic batch method of concentrating material, has in many cases been replaced by vacuum evaporators (Fig. 21). Typical products are concentrated (frozen) orange juice, other fruits like apricots and peaches, milk, tomato juice, sauce, and paste. Some products lend themselves to concentrations as high as 50% solid without quality depreciation.

Types of units used range from single to multiple effects and, depending largely on product characteristics, utilize single-pass or recirculation flow principles with high temperatures (ordinarily steam) or low temperatures (provided by a heat pump). Efficient utilization of the heat pump is achieved by using heat from the compressor's cooling water to heat the product and refrigeration to chill water for the barometric condenser. Some units are equipped with fractionating columns to recover product essence, often lost in discharged vapors, which subsequently is combined with evaporated material to recapture its natural flavor. Some plant wastes, for example, whey from cheese manufacture, or sardine stick-water, now are recovered by evaporation and used as feed. All modern evaporation units are highly automated by instrumentation to conserve labor costs and assure uniform quality of end products. All product-contact surfaces are of stainless steel.

Drying and dehydration is done either in a batch-type or continuous operation. Batch-type drying using cabinets or trays on mobile racks that are moved through tunnels by chain or hydraulic ram are typical but less efficient than continuous belt-type dryers. The continuous type unit employs through-product, heated air circulation. It provides means for better control of moisture-humidity; permits deeper product beds and thus increases production capacity while conserving labor and steam.

High ratio of heating surface to product volume is achieved with the drum dryer. A thin layer of product is spread on the surface of an internally heated, rotating hollow cylinder or pair of cylinders. The product, dried to a 1–4% moisture in one revolution, is then removed from drum by a scraper or "doctor knife."

A belt-through dehydrator, originally designed for dehydration of food particles as part of the dehydrofreezing process, is also used as a dryer (Fig. 22). Its inclined, moving belt imparts a fold- [illegible] achieves [illegible] to con- [illegible] milk to a readily reconstituted dry powder. Materials such as solutions or slurries are introduced into the drying chamber either by high-speed centrifuges (6000–20,000 rpm) or high-pressure nozzles (2000–10,000 psi) into a heated atmosphere (137.8–148.9°C) moving at high velocity. Evaporation from the atomized mist takes a fraction of a second, with over-all drying time usually under 30 sec. The largest volume of dried product (1-3½% moisture) falls to bottom of chamber and is continuously removed. The fine particles, entrained in exhaust air, are recovered by cyclone collectors.

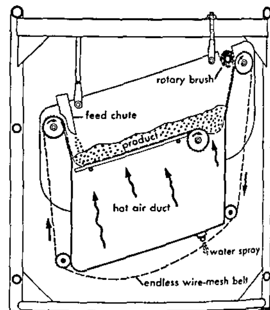


Fig. 22. Line drawing of belt-trough dryer. A spiral movement is imparted to product to gain drying efficiency.

Continuous vacuum dehydrating occurs in a vacuum chamber, where the concentrate (for example, orange juice of 58% total solids) is transferred from a feed trough, by a roller, to an endless, stainless belt, powered by two drumlike pulleys. One drum heats the belt, the other cools it. Complete dehydration takes place as the film-covered belt moves from the feed roller across the heating drum to the cooling drum, where a doctor knife shears off the crystalline powder.

Freeze-drying is ordinarily done in a vacuum cabinet with heating and cooling facilities. This process has been used for over 50 years in the biological sciences for preservation of unstable materials. Fundamentally, it consists of water removal under high vacuum by sublimation from the solid phase, ice, to the gaseous phase, water vapor. Various types of cabinets or refrigerated ovens are used. All produce high vacuums of 300–700 millimicrons. With some cabinets, evaporation takes place directly from the solid phase without passing through the liquid phase. Materials dried by this method range from semisolids for meats to powders for juice concentrates.

Dehydrofreezing is a dual process requiring both dehydration and freezing. A quality product of light weight is produced. Pimientos were used in the first commercial application of the process, and were followed by peas. [C.B.H.]

Irradiation. This is one of the newest methods of preserving foods. The method of conversion of raw materials by irradiation to extend their storage period commonly refers to the use of ionizing radiations.

In this category are the penetrating radiations like γ- and x-rays, and those less penetrating radiations such as cathode rays or β-particles. *See* CATHODE RAYS; GAMMA RAYS; X-RAY(S), PHYSICAL NATURE OF.

Gamma rays arise from the nuclei of radioactive isotopes during their decay processes, whereas x-rays are produced by man-made accelerators. Both have the same properties at the same energy levels.

Cathode rays, electrons, or β-particles are synonymous in terms of physical characteristics and effects. Cathode rays are streams of electrons accelerated by a high-voltage potential. Beta particles are actually electrons emitted from the nuclei of radioactive atoms during their decay processes. For uses in food and drug sterilization, energy levels above 1,000,000 volts are needed.

These ionizing radiations have been used for extension of storage life or pasteurization, and for sterilization in recent experiments.

Radiopasteurization may be accomplished by surface treatment, in which case low-energy (under 1 Mev) electrons are used. An example is the extension of the storage life of frankfurters by reducing the mold and bacterial formation on the surface.

Reduction of the total microbial flora throughout a product by substerilization doses of 1,000,000 rad or less describes the process of penetrating radiopasteurization. An example is the extension of the storage life of hamburger by treatment with doses of 500,000–1,000,000 rad, thereby reducing the bacterial load from 10^6 bacteria per gram to less than 10^3 bacteria per gram with resultant increase in storage life at 7.2°C from 1 week to 1 year.

Sterilization refers to the utilization of doses high enough to result in the conventional type of commercial sterilization considered in terms of thermal processing. In this case, depending on the type of spoilage organisms, doses of 10^5 to 5×10^6 rad are required.

Antibiotic treatment. This has been used since 1953 to extend the storage life of certain foods under refrigeration. Certain of the antibiotics, such as tetracyclines, have been approved by the U.S. Food and Drug Administration for use in dilute solutions into which eviscerated poultry is dipped. In the United States, Canada, and other countries dilute solutions can be used to make ice for refrigerating fish.

Fermentation. Fermentation is a means of preserving foods or developing new foods by the conversion of various carbohydrate materials through the use of microorganisms. Examples of fermentation in the production of foods are the following: the conversion of lactose to lactic acid with the resultant end product of fermented milk, such as yoghurt or the resultant end product of cottage cheese or other cheeses; the conversion of flour by yeasts to break down the sugars into carbon dioxide, thus producing a fluffy loaf of bread; the conversion of sugars in cabbage and cucumbers to produce sauerkraut and pickles. *See* CHEESE; FERMENTATION; INDUSTRIAL MICROBIOLOGY.

In the Far East, a number of soybean products are produced by fermentation processes. An example of this is the Japanese product *miso.* This is produced by the action of *Aspergillus,* a mold, on cooked soybean.

The fermentative processes take place through enzyme reactions. The enzymes, produced by the microorganisms in their growth and metabolism, in turn convert the various raw materials cited above.

Pickling, curing, and aging. The processes of pickling, curing, and aging refer mainly to applications in the meat industry and to a lesser extent in the fish and dairy industries. Various meats such as pork shoulders, hams, and briskets of beef are soaked in brine solutions containing sugar and spices. Included in the pickling solution are sodium nitrite and sodium nitrate which produce the typical pink coloration of pickled meats.

Following the pickling process some of the meats, such as hams, may be smoked. Others like corned beef are not smoked.

The addition of pickle results in longer shelf life due to the increase in salt in the food product.

Curing generally refers to the storage of products like cheddar cheese for a period of time under controlled environmental conditions in order to develop taste, aroma, and proper texture. The changes that take place in the curing process are a result of the enzymes from the microorganisms. The temperatures for curing usually range from 0–4.4°C and the time required is from 3–12 months.

Aging refers generally to cheese and other food products and in particular to alcoholic beverages which are also one of the most important industrial examples of fermentations.

During the aging process various harsh-flavored materials, such as the fusel oils, are converted to more mild flavored materials. The aging process may continue for a period of a few months to several years. During this time, changes take place in the solids, esters, acids, and aldehydes, as well as fusel oil, with a resultant better flavored product. *See* ETHYL ALCOHOL; MALT BEVERAGE. [S.A G.]

Smoking. Smoking is a method for enhancing flavor and extending shelf life of products. This is accomplished by volatilization and redeposition of certain components of hardwoods in smokehouses The smokehouses are generally built of sanitary stainless steel with excellent temperature and humidity control (Fig. 23). The substances volatilized and redeposited are pleasantly flavored in addition to possessing bacteriostatic properties.

Although many methods of smoke application have been used, the ones of prime importance are burning of hardwood sawdust; destructive distil-



Fig. 23. Modern 4-cage smokehouses which are provided with automatic temperature and humidity controls.

tion of smoke particles. One recent innovation consists of vaporization of liquid wood distillate.

Time, temperature, and relative humidity are carefully controlled during the smoking process; these conditions are varied according to the type of product being smoked.

Churning. Churning is a mechanical mixing process used essentially to separate the fat phase from a fat-water system. The process is universally used in the manufacture of butter. Churning and modifications of churning are also employed in the production of margarine.

The first step of a churning operation converts an oil-in-water emulsion (O/W) to a water-in-oil emulsion (W/O) through agitation and the incorporation of air. Once this conversion has been attained, with continued mixing the fat will agglomerate from the aqueous phase. After removal of the aqueous phase, churning is continued to incorporate salt and moisture and to standardize the composition of the butter.

Churning is a batch process carried out in horizontal rotating cylindrical vessels normally designed to produce approximately 1000 lb of product. Churns are constructed of wood or metal and contain baffles or vanes for improving agitation and working of the contents.

Continuous methods for making butter which replace churning have found some acceptance. *See* BUTTER; MARGARINE.

Deodorizing. Deodorizing is a steam distillation process for the removal of small amounts of volatile odor and flavor substances (Fig. 24). The process is routinely used in the preparation of fats and oils for the manufacture of shortenings and margarine. Off flavors are occasionally removed from cream by deodorization.

Deodorizing is carried out in batch and in continuous equipment. In both, the material to be deodorized is heated under reduced pressure and treated with steam to strip out the volatile components.

Plasticizing. Plasticizing is a mechanical process employing rapid chilling, agitation, and gas incorporation for improving the texture and consistency of a fat or fat product.

Most fats, when cooled slowly from the liquid state, form large crystals and produce a grainy product. When they are cooled rapidly, the crystal size is small and the texture is smooth. Additional working of the chilled fat and the incorporation of

[illegible] ctional proper-

[illegible] the shortening,

[illegible] creaming, and volume-producing characteristics of fats used in baked products and the spreadability of margarine.

Plasticizing is carried out continuously in chill roll equipment or in closed systems employing concentric-tube heat exchangers equipped with agitation and gas incorporation devices.

Hydrogenating. This is a chemical process for converting liquid fats into solid or semisolid fats. It is a catalytic process which adds hydrogen to ethylenic or double bonds in the fatty acid chains of the fat. *See* HYDROGENATION.

Hydrogenated fats have higher melting points and improved plasticity. They are more suitable for the manufacture of shortenings and margarine. Hydrogenation also removes undesirable odors and colors from fats and increases their stability.

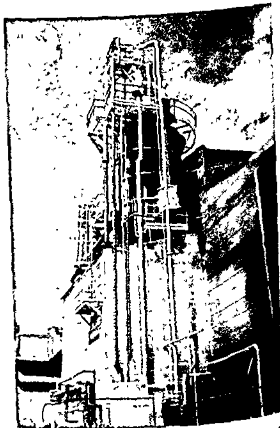


Fig. 24. Continuous vegetable oil-deodorizing unit (Blaw-Knox Co.)

Hydrogenating is normally carried out batchwise in 10- to 20-ton capacity agitated converters equipped for heating and cooling. Refined oil, containing a dispersed catalyst (usually nickel), is heated to reaction temperature and hydrogen is ad-

...

ceeded to the extent desired, the fat is cooled, filtered to remove the catalyst, and deodorized.

The properties of the hydrogenated fat can be ... which the re-

...

temperature, pressure, the type of catalyst used, and the concentration of the catalyst.

Decolorizing. Decolorizing is a mechanical separation process for the removal of naturally occurring undesirable color bodies. Decolorizing or bleaching is an adsorption separation in which the impurities are selectively adsorbed on an activated surface.

The process is used primarily in the refining of fats, oils, and sugar. Adsorbents used are normally bleaching clays or earths and carbon.

Decolorizing is performed in batch and continuous equipment. A filtration step follows the decolorizing process to remove the adsorbent.

[C.K.W.]

Puffing. A process that expands foods to several times their original volumes is called puffing. It is a key operation in converting cereal grains and cooked doughs into palatable dry breakfast foods, such as puffed wheat and rice.

Moist grains are heated (with or without added steam) in an enclosed, sanitary pressure chamber where, almost instantly, accumulated pressure is released by opening a large, quick-acting valve (Fig. 25). Resultant expansion of vapor within the grain or dough explodes the product. Process variables such as product moisture content, pressure, and temperature must be carefully regulated.

Powdered citrus juices and other dehydrates are also puffed, in this case during vacuum-drying, to increase their solubilities.

Whipping. Whipping incorporates gases (usually air) into relatively viscous liquids, and is essential in the production of ice cream, mayonnaise, and marshmallow. Agitators employed for whipping are of three general types: (1) beaters that fold gas into the liquid; (2) shaft-mounted propellers that rapidly draw liquid between the blades; and (3) toothed disks that rotate between toothed stators. The last two units cause churning.



Fig 25. Vacuum puffing unit. Vacuum puffer expands confections, such as mints and malted centers, as much as 30 times their original volumes, while evaporating 1–3% of product moisture. At feed end (left) are seen unit's double vacuum locks. (*F. J. Stokes Machine Co.*)

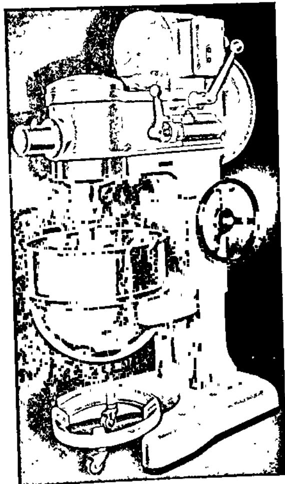


Fig. 26. Bakery whipping unit. Beater has sanitary, open-rim bowl that fits on 4-caster dolly. (*Hobart Manufacturing Co.*)

In the manufacture of ice cream, the mix is frozen while a definite volume of air, called the overrun, is incorporated by a specially designed ... is whipped in the rotating-disk ... air to mini- ... (Fig. 26).

Deaeration. Deaeration is the removal of air from foods and beverages either to curb oxygen-induced spoilage or to reduce the volume of a powdery product.

Bulk liquids are deaerated by being subjected to vacuum, usually with application of heat. Packaged carbonated beverages like beer are deaerated by striking the container. This causes foaming that eliminates the air in the headspace.

Powders are deaerated by vacuum processing; the evacuated air may be replaced by an inert gas.

such as nitrogen. This practice greatly extends the keeping quality of products like dried whole milk, soluble coffee, and powdered fruit juices.

Crystallization. Crystallization is the separation of relatively pure solute from a saturated solution as it is evaporated or cooled. Since the precipitated crystals may entrain some of the impurities of the mother liquor, they generally are centrifuged and recrystallized from pure solvent to obtain a pure product

A major refining step in the processing of sugar, salt, and monosodium glutamate, crystallization is usually carried out in vacuum pans which are large single-effect evaporators. It may occur spontaneously as the solution becomes progressively supersaturated, or can be induced by seeding the boiling solution. Precise control of evaporator temperature, vacuum, and feed rate is essential for maximum yield of proper-sized crystals

Emulsifying and homogenizing. These are techniques for producing substantially nonseparating mixtures of insoluble materials by mixing and particle-size reduction. Typical emulsions and suspensions thus manufactured include dairy products, mayonnaise, and fruit juices.

Homogenizers exert intense shearing action on the particles of a mixture as they are pumped at high pressure through minute orifices. A similar effect is achieved with colloid mills, which employ closely spaced high-speed rotors and stators (Fig. 27).

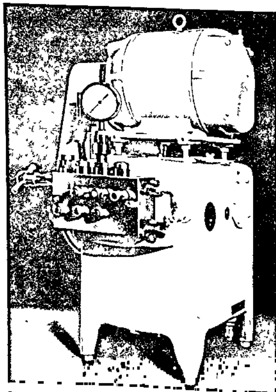


Fig. 27. Homogenizer. Employed is a 3-cylinder pump having a minute adjustable orifice through which product is forced under high pressure. (Cherry-Burrell Corp)

Ultrasonics offers another rapid and efficient means of stabilizing food mixtures. Vibrations with frequencies in excess of 18,000 cps are transmitted to the product. Ultrasonically homogenized peanut butter, for example, is now being produced.

Coatings. Coatings are generally applied to baked goods, ice cream novelties, and confections to enhance the taste and the attractiveness of these products.

Rectangular and round cakes are iced by machine, which deposits the frosting in a sheet. Sweet rolls, buns, and Danish pastry are iced by high-speed continuous units as they are conveyed under print rollers that give the icing a hand-finished appearance.

Confections such as hard candies, nuts, and chewing gum are sugar-coated in heated pans that slowly rotate at an angle of about 60° to the horizontal. Successive portions of supersaturated sugar solution, containing flavoring and color (if required), are added and evaporated from the product surface. This builds up the coating by depositing thin layers of sugar.

Chocolate-coated products, such as ice cream and candies, generally are coated by enrobers. However, hand-dipping is still practiced. In machine operation, products are first bottom-coated as they are conveyed (on a wire belt) over a shallow layer of molten chocolate. For top coating, they pass through curtains of molten chocolate that cover them completely (Fig. 28). Finally, the coating is set in a cooling tunnel. [L.T.]

Food packaging. This is a multiple unit operation involving machine filling of foods into rigid or nonrigid packages that are sealed, labeled or overwrapped, coded, and then packed into shipping containers.

The basic problem is to meet a product's protective, merchandising, and economic demands. A prime objective is greater efficiency by reducing unit package costs through proper selection of newer, lower-cost, protective packaging materials as well as through rapid development of better, faster, mechanized packaging-line operations. While these developments are improving food packaging, they are also making it more complicated. The machines and operating lines are more complex and automatic by comparison with those in use previously.

Packaging materials. Depending upon the commodity, the package can be a can, bottle, jar, carton, bag, envelope, pouch, or wrapper. In addition, it can be a composite like a bag or inner liner within a carton, a bottle within a carton, or a wrapper over a carton. The package will employ in its make-up such materials as metal, glass, paper, plastic, adhesive, ink, and coating.

The sanitary can. The sanitary can is made of tin-coated steelplate, which is flanged at both [illegible] ...led-rim end [illegible] composi- [illegible] nd rims are machine-folded together, the rubber composition provides a tight, gasketlike seal. The body is



Fig 28. Enrober. Enrobing unit receives candy bars on plastic conveyer belt, coats them with tempered chocolate. The chocolate is piped to machine from melting kettles. (National Equipment Co.)

delivered to the food packer with one end sealed; the other end is sealed when the can is filled.

Plain tinplate cans are suitable for many products, but organic coatings for can linings are employed when additional protection is needed. Linings also increase attractiveness of foods and improve container appearance.

About 50% of the cans produced are used for foods. Of the 32 sizes recommended, only 9 are widely used by food canners.

Top can-filling and closing-line speeds run up to 1000 per min. Average speeds are about 250 cans per min.

Glass containers. About 40% of the glass containers produced are used for foods. Glass is suitable for foods because it can be molded into various shapes and is inert and transparent.

Glass is molded into containers of innumerable shapes and sizes, large or small, tall or squat, and narrow- or wide-mouthed, and used for liquids, solids, and semisolids. Because glass is inert, it does not add anything to or take anything from a product, and due to its transparency, glass displays foods in their natural color and shape.

Modern machines turn out glass containers much faster than old-time blowers, and also make them lighter, stronger, more suitable, and more uniform in color. Improvements in closures and production methods have also increased their popularity. Effective, convenient closures and improved containers have increased packaging speeds, resulting in better-quality products as well as an increased number that are packed in glass (Fig. 29).

Line speeds ran slightly under 100 containers per minute in 1935. In 1959, speeds of 200–350 containers per minute were common. Baby-food jars were being run off at over 500 per min.

Cork was once almost the only closure used for bottles. Numerous types of gaskets, suitable for a variety of products, have been developed by closure manufacturers. With proper gaskets and high-speed machine application, lug and screw caps are suitable for almost all products.

Standardization of container finishes in 1930 sparked growth in the use of glass. A packer can purchase caps and glass in confidence from various sources by specifying the Glass Container Manufacturers' Institute finish.

Closures for glass containers are mostly made of metal or plastic, though paper is used in the dairy industry for glass and paper containers to hold liquids. Closures prevent escape of contents and entrance of foreign matter. The seal should be easily broken by the customer, while the closure should withstand several reclosings.

Major closure types include crown, vacuum, rolled-on, screw cap, lug cap, tamperproof, snap-fit, and press-on.

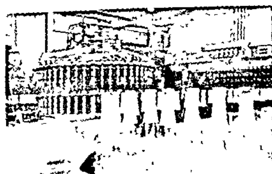
Paper. This basic packaging material is modified or combined with other materials to offer many product-protection features. Among other properties, it is designed to be impervious to gas, grease, water, or moisture vapor; to resist insect penetration, corrosion, and mold growth; and to withstand impact, tear, puncture, water penetration, and embrittlement. It is also made nontoxic, odorless, and tasteless. *See* PAPER AND PAPER PRODUCTS.

Protective papers include (1) plain kraft (from sulfate pulp), (2) laminated (asphaltic) kraft, (3) coated (heat-sealable polyethylene) kraft, (4) waxed paper (thin-calendered wax-coated paper), (5) wet-strength (added resin) paper, (6) vegetable parchment (acid-treated cellulose), (7) greaseproof paper (dense, uncalendered paper from hydrated pulp), and (8) glassine (dense, supercalendered paper from highly refined special pulp).

Glassines may be plain lacquered (coated with heat-sealable, resinous materials), wax-coated, or laminated (several plies of glassine with wax laminant).

Paperboard containers. These comprise liquidtight containers, paper cans, set-up boxes, and folding boxes.

Lightweight, liquidtight containers are essentially used for dairy products, frozen foods, and delicatessen items. They are of either the nested type, made of virgin sulfate or sulfite stock and treated with paraffin or untreated; or of canister shape (conical, cylindrical or rectangular), of manila stock, plain or lined with sulfite or sulfate. For specific uses canisters may be glassine-, cellophane- or parchment-lined, and they may be



29. Beer filling line. (Miller Brewing Co.)

treated with paraffin or vinyl formulas or untreated.

The paper or fiber can is a rigid, multi-ply paper body with metal or paper ends. The all-paper type is cylindrical, while the composite type, made of a paper body with metal ends, is round, square, oval, oblong, and rectangular.

The body wall is generally fabricated from chipboards and box boards, while combinations of these and kraft, jute, foil, glassine, and others may also be employed. A wide range of adhesives is used for multi-ply-constructed bodies that are usually wax- or paraffin-coated on the inside. Paper ends are generally of strawboard with a colored liner outside used plain or in combination with inside foil, glassine, or parchment. Tinplate or lacquered blackplate make up construction of the metal ends.

Set-up boxes are of a rigid type, made from paperboard and delivered ready for use. Generally, they consist of top and bottom, but they vary widely in type and shape. They are used for confections, baked goods, and certain prestige items to create impulse buying and protect products in storage and distribution.

There is an almost endless variety of folding-carton structural styles However, carton constructions are classified as tray or tube type.

The tray type is an unbroken paperboard bottom panel, with each side and end wall panel connected to its adjacent wall by a glue flap, a hook-engaging slit in the wall or some other connection. These may be of two-piece construction—base and cover with or without transparent window.

The tube type contains a seam or flap glued by the carton maker. Styles include the seal-end carton with one end closed, the tuck-end construction with one end closed, a combination tuck-end and lock-end closure, and a lock-bottom construction.

Aluminum foil. Plain unsupported foil is used as semirigid containers for specialty foods, frozen cooked foods, and bakery products; as wraps for candies and cheese; and as overwraps for cartons and frozen-food trays.

Aluminum foil in combination with paper and plastic films and heat-seal coatings is employed as carton overwraps, direct wraps, shipping case liners, labels, bags, pouches, brown-and-serve trays, ice cream containers, carton liners, and canister liners. Thickness ranges from 0 00025 to 0.006 in.

Cellophane. This transparent, versatile film consists of a plasticized base sheet (regenerated cellulose), which is essentially greaseproof, and can be made moisture- and vaporproof through coatings. Its functional properties are [illegible] type for almost every conceivable application. Basically, it comes in four forms: uncoated, intermediate moistureproof, moistureproof, and water resistant. Thickness ranges from 0 0008 to 0.0017 in.

Plastic films. These are made from synthetic resins or polymers or natural polymers, and nearly all are sealable by heat and pressure. They can be modified for certain applications, and possess many types of protective properties in addition to being attractive and highly durable. Basically, they include the following types.

Cellulose acetate differs from cellophane in that the acetate base is water insensitive and soluble in certain solvents. It is used for packaging fruits and vegetables because of insensitivity to water softening and high rate of gas and water-vapor transmission; for window material in cartons; and for rigid transparent containers.

Polyethylene is characterized by extreme toughness and rubberiness. It is formed into film by casting from solvent solution, extruding, or calendering. Polyethylene is finding increasing applications because of its excellent physical and protective qualities and relatively low cost.

Polyethylene resin is used for coating paper, paperboard, and various films and is used as one of the plies in a range of laminated combinations (for example, extrusion-coating it directly onto cellophane).

Pliofilm is a rubber-hydrochloride, heat-sealable film that is available in some dozen basic types and in eight grades. Various types are modified with stabilizers and plasticizers to produce films that excel as a barrier to gas and water vapor, are tough and flexible, and serve for heavy-duty special purposes.

Vinyls are resins of vinyl chloride or blends of vinyl chloride, vinyl acetate, and other resins. The film's use in the food industry is currently limited, but there is future potential.

Saran is an extruded vinylidene-chloride film with such outstanding properties as strength, cling, and high temperature resistance. It has some significant limitations, however, since problems are encountered in heat sealing and machine handling.

Polyesters form an exceptionally tough film made from polymer of ethylene glycol and terephthalic acid. The film is finding use as transparent window material for cartons, bags, and wraps. Its initial success was as a laminant for cook-and-serve packages.

Laminated materials. Two or more plies of similar or dissimilar materials are made into a single sheet. An endless variety of laminated materials are tailor-made to meet requirements of specific packaging problems. Selection of materials influences properties desired, like uniting decorative appeal of aluminum foil with the mechanical strength and greaseproof quality of glassine to produce an excellent candy wrapper.

A large number of rigid plastic containers are made in molds. Those made with styrene and cellulose acetate are principally used for foods. *See* PLASTICS FABRICATION.

New specialized adhesives and related materials have shown the way to wider use of films, foils, and papers. Recent advances in adhesives have improved production and performance of containers. Development of an adhesion promoter, coated or pattern-printed onto polyethylene extrusion-coated

board or paper, has permitted the use of versatile polyvinyl acetate emulsion adhesives.

A wide variety of rubbers and rubberlike resinous materials are used for pressure-sensitive adhesive applications such as tapes and labels.

Generally, adhesives are compounded from starch or its derivatives, casein, animal glue, lacquers, latex, resin emulsion, and heat-sealing materials (hot melts, emulsions, and solvents).

Packaging methods. The selection of machinery and packaging lines is influenced by the type of product and package handled. Packages are fed to the line for product filling, sealing, labeling or overwrapping, coding, and packing into shipping cases that are sealed.

Feeding packaging lines. Cans and glass containers are automatically unloaded from shippers and dumped in helter-skelter fashion onto an unscrambler for single-file delivery onto the line, right side up, for filling.

The unloaders handle 15 or more cartons per minute. One type conveys cartons upside down as containers drop onto the unscrambler after carton flaps are opened. Another employs vacuum cups to engage the bottoms of bottles (packed upside down), lift them from cartons, and invert them for feeding to the unscrambler. Still another uses mechanical fingers to grip and unload narrow-neck bottles, and still another unloads metal cans by electromagnetic pick-up.

Unscramblers of the rotary and straightline types employ disks, belts, gates, and reverse-flow principles. A noiseless, gentle-handling, walking-beam type unscrambles 500 containers per minute.

Accumulators serve as a temporary storage area for the steady flow of containers to the packaging line in the event of shutdowns or variations in line-machine speeds. One type employs a revolving disk, another a trough or conveyor-belt-fed table area.

Bottle cleaners for new bottles are of water rinsing and air cleaning types. One automatic rinser feeds bottles into carrier-mounted pockets, alternately subjecting them to air jets, water-conditioning spray, prerinse, draining, and right side transfer onto conveyor.

An air cleaner receives bottles from the conveyor and injects air while the star-wheel discharges them onto conveyor.

Filling machines. Fillers are designed for dry, liquid, and viscous products.

Dry fillers have speeds up to 350 cartons per minute, and employ carton filling and sealing machines for free-flowing products. The unit has a vibrator, a check weight scale, and electronic controls to adjust weight variations.

Filling lines are appearing in which bags are opened, positioned, filled (by weight), and sealed. Basic types of filling machines include (1) net weighers to scale the product and fill the weighed charge into the package (Fig. 30); (2) gross weighers to scale the product and package; (3) volumetric fillers to measure out the product with a rotating plate containing calibrated pockets; (4) metering fillers to establish a standard rate of product flow for filling into containers at specific time intervals; and (5) auger-type feeders to deliver the measured amount of product.

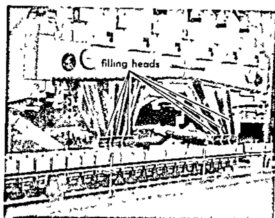


Fig 30. Six-head carton filler (Pneumatic Scale Corp., Ltd.)

Liquid filler operations are speeded through improvements in machine design, and by smooth and efficient handling of metal and glass containers. Examples of production rates per min are 350 soft drink bottles (12 oz), 700 bottles of ketchup (14 oz), 800 cans of juice (6 oz), and over 1000 cans of baby food.

Liquid fillers for containers with mouth openings below 38 mm cover those operating on the principle of (1) gravity filling, liquid flowing by gravity from an overhead surge tank through filling valves, and cutting off at a predetermined level in the container; (2) gravity vacuum, gravity filling in a closed-vacuum system; (3) pressure gravity, using a booster pump to produce pressures 10–20 psi to increase speed of gravity filling of heavy viscous liquids; (4) vacuum, evacuating air from container to force liquid into it from a reservoir below; (5) pressure-vacuum, employing modification of standard vacuum filler with pressure feeder; and (6) volumetric, utilizing a filler tank with an accurately controlled liquid level, and measuring cups to deposit a premeasured amount into the container.

Viscous fillers in general use are of two basic types, straightline and rotary plunger. The difference is that straightline fillers operate intermittently on a given number of containers, while rotary plunger fillers fill and discharge containers continuously. Both employ plungers for accurately measuring and forcing the product into containers.

Closing machines. The equipment employed is influenced by the type of package being handled. Bags or pouches containing thermoplastic material are generally sealed by application of heat and pressure. Heat sealers include roller, band, jaw, high-frequency, or impulse types for sealing unsupported films, coated materials, laminations, and combinations of coated and laminated materials. Cartons are conveyed from the filling machine past the glue applicator, flap closer, and compression belt. Cans are sealed by rotary seamers that apply and crimp the top lid to the can body.

For sealing glass containers, rotary or straight-line automatic capping machines are employed to handle lug-type caps, side-seal closures, and continuous thread caps.

Labeling machines. Labels are applied by machine to cans, package overwraps, jars, or bottles, either from roll-fed stock or from stacks. Cans are horizontally fed through a wrap-around labeler, where a moving belt carries them past the glue applicator and under the horizontal label stack. Pressure-sensitive labels are applied from roll-fed stock onto overwrapped packages

jars. Machines range from simple table model hand-operated gummers to fully automatic units capable of bringing together, in an assembly operation, the container and various combinations of labels numbering as many as four to a container. Plain labels are glue-applied and thermoplastic labels are heat-applied to the containers. They are applied to the front or simultaneously to the front and back body of the containers, as well as to the neck and body of bottles. Speeds run up to 300 jars per minute

Wrapping machines. These perform their operations in many ways. In one machine, the package is lifted by a reciprocating elevator through the time sequence of positioning and wrapping. Lengthwise wrapper overlap is sealed against the bottom of the package, with end folds either turned under the package and sealed against the bottom or folded and sealed against package end panels.

Other wrappers achieve similar results (without an elevator) through the use of intermittently rotating pockets, where several folding operations are performed. In others, the package is carried through all operations in a horizontal plane and past an L-shaped folding line that wraps, folds, and seals two ends and one edge of the package.

Typical speeds are 180 cartons of frozen food and 1200 chewing gum sticks per minute.

Coding devices. Packages and shipping containers carry markings indicating the code date, batch number, lot number and other identification legends. Markings are machine-applied to packages by inking, notching, perforating, indenting, or embossing. Imprinting attachments are employed for wrapping, bag-making, and bundling machines which handle material in roll form.

Coding, marking, and imprinting machines are available as attachments for other packaging or conveying equipment; there are also in-line units that fit along with machines for filling, capping, sealing, cartoning, and casing, as well as independent units.

Multipackaging machines. These units are a complete, automatic packaging line. They handle roll-fed, sealable material, which is formed into pouches, filled with liquid, dry, or viscous products, and then sealed. Some machines print and code pouch material or insert one, two, or three pouches into cartons. Others seal pouches by vacuum or inert gas. Speeds are up to 3000 or more packages per hour. Feed mechanisms include scale, volumetric, auger, pocket delivery, count, or liquid filling.

Casing and bundling machines are used to accumulate cans, cartons, and glass containers automatically for semiautomatic or automatic loading into shipping cartons which are then glue-, tape-, or staple-sealed. Case loaders include (1) gravity roll-in types in which cans are tiered (by rolling) to a proper pattern as the operator places the shipping container on a loading horn, releasing a plunger to insert the cans into a container; (2) offset-conveyor-feed type, in which cartons are elevated and grouped by an offset infeed conveyor; the container is manually placed on a loading horn and loaded with accumulated cartons; (3) elevator-feed type, in which cartons are fed into the case loader via the single- or multiple-lane in-feed belt and mechanically elevated until a prescribed pattern is formed; and (4) case feeder, former, positioner, loader, and sealer machine collects margarine cartons from ten wrappers, and accumulates groups, and loads 30 cartons into an end-loading, preformed, shipping container.

In the bundling operation, packages are conveyed from wrapping or carton filling-sealing machines to the bundler, which assembles them into units of predetermined count and pattern. The bundle is then machine-wrapped in paper, film, paperboard, or corrugated board. The bundler, using glue or heat, may apply separate end seals from roll or magazine for identification or appearance, may imprint wrappers with a mounted imprinter, and may apply easy-opening devices.

[J.V.Z]

Bibliography: Am. Meat Inst., Committee on Recording, *Pork Operations*, 5th rev. ed., 1944; Am. Meat Inst., Committee on Textbooks, *Beef, Veal and Lamb Operations*, 4th rev. ed , 1945; Am. Meat Inst., Committee on Textbooks, *By-Products of the Meat Packing Industry*, rev. ed., 1950; American Bottlers of Carbonated Beverages, Bulletins and Reports, 1957–58; A. E. Bailey, *Industrial Oil and Fat Products*, 2d ed., 1951; C. O. Ball and F. C. W. Olson, *Sterilization in Food Technology*, 1957; F. C. Blanck, *Handbook of Food and Agriculture*, 1955; C. Butler, J. W. Slavin, and F. B. Sanford (eds.), *Refrigeration of Fish*, pts. 1–5, U.S. Fish Wildlife Serv., Fishery Leaflet 427–431, 1956; A. T. Cameron, *The Taste Sense and the Relative Sweetness of Sugars and other Sweet Substances*, Sugar Research Foundation, N.Y. Sci. Rt. Ser. 9, 1947; R. J. Clarke, *Process Engineering in The Food Industries*, 1957; W. V. Cruess, *Commercial Fruit and Vegetable Products*, 4th ed., 1958; A. C. Dahlberg and E. S. Penczek, *The Relative Sweetness of Sugars as Affected by Concentration*, N.Y. State Agr. Expt. Sta., Bull 258, 1941; S. S. Epstein, Beverage quality, *Food Eng.*, 23(9): 100–102, 1951; Food Eng. staff, Diffusion problem whipped, *Food Eng.*, 25(2):56–57, 223–224, 1953; O. F. Hunziker, *The Butter Industry*, 3d ed., 1940;

N. D. Jarvis, *Curing of Fishery Products*, U.S. Fish Wildlife Serv., Research Rept. 18, 1950; N. D. Jarvis, *Principles and Methods in the Canning of Fishery Products*, U.S. Fish Wildlife Serv. Research Rept. 7, 1943; M. B. Jacobs, *The Chemistry and Technology of Food and Food Products*, 2d ed., 1951; S. Jordon, *Confectionery Problems*, 1930; E. W. Meeker, Confectionery sweetners, *Food Technol.*, 4(9):361–365, 1950; M. E. Parker, E. H. Harvey, and E. S. Stateler, *Elements of Food Engineering*, vol. 2, 1954; S. C. Prescott and B. E. Proctor, *Food Technology*, 1937; B. E. Proctor and S. A. Goldblith, Present-day status of radiation sterilization, in *Atomic Energy and Agriculture*, Am. Assoc. Advance. Sci. Publ. 49, 1957; B. E. Proctor and S. A. Goldblith, Preservation of foods by irradiation, *Am. J. Public Health*, 47:439–445, 1957; G. F. Reddish (ed.), *Antiseptics, Disinfectants, Fungicides, and Chemical and Physical Sterilization*, 2d ed., 1957; M. K. Schwitzer, *Margarine and Other Food Fats*, 1956; N. O. Simmons, *Compound Milling*, 1955; D. L. Traux (ed.), *Sanitation for the Food-Preservation Industries*, 1952; D. K. Tressler and C. F. Evers, *The Freezing Preservation of Foods*, vol. 1, 3d ed., 1957; D. K. Tressler and M. W. Lemon, *Marine Products of Commerce*, 2d ed., 1951; H. W. Von Loesecke, *Drying and Dehydration of Foods*, 2d ed., 1955.

Food manufacturing

The commercial production and packaging of food products that are fabricated by processing or combining various ingredients or both. Manufactured foods are basically different from any found in nature, but they are fabricated principally from natural ingredients, and often include small quantities of chemical food additives to improve nutritive value, taste, color, shelf life, and convenience in use.

Mechanical manipulation, chemical treatment, biologic processes, heat treatment, or freezing may be involved. For example, bread is manufactured by combining flour, shortening, yeast, and other minor ingredients to make dough. Kneading and fermentation of the dough change the characteristics and develop flavor. The dough is formed into loaf-sized pieces, put into bake pans, proofed, and baked to form bread, which is sliced and wrapped.

Some foods, however, are prepared for consumption by refining processes which do not essentially change the nature of the material. Milk, for example, is clarified by centrifuging and pasteurized by heat treatment to kill pathogenic organisms. This treatment makes it safe for human consumption, but does not change it basically.

Food manufacturing and processing involve many unit operations, such as separating, mixing, forming, application of heat or refrigeration, packaging, materials handling, and process control. These are carried out in some 40,000 plants in the United States, ranging in size from small retail bakeries to a canning plant covering 14 acres. The food industry has 1,640,000 employees, more than any other industry. It spends some $800,000,000 a year for modernization and expansion, purchases huge quantities of raw materials from farmers and ingredient manufacturers, uses some $7,000,000,000 of packaging materials and containers, and produces more than $50,000,000,000 a year of finished foods for homes, mass-feeding institutions, and export.

Originally an enlarged version of a kitchen, the typical food factory has been developed through engineering into a highly efficient, technical industrial plant. Many processes have been converted from batch to continuous methods and put under automatic control. Materials and finished products are handled with high efficiency and little labor, with the aid of bulk pneumatic conveyors, pumps, mechanical conveyors, and lift trucks. Modern food factories are built with sanitary floors and walls of tile and brick, and much of the equipment is made of stainless steel. Most of the newer plants are in suburban areas on attractively landscaped plots. Aside from being built with sanitation in mind, many modern plants are showplaces through which visitors are taken on regular tours as a public relations operation.

Food factories are inspected by Federal health authorities if they do interstate business, and by state and municipal authorities when they do business locally. Health agencies also require that additives and packaging materials used in food manufacturing be approved as safe before they may be used.

Manufacturers usually sell their products to retail stores through wholesalers, though some sales are made direct to large chain-store organizations. Most dairies and many bakeries also deliver directly to the home.

Major branches of the food manufacturing and refining industry are those producing bakery products; flour and cereals; milk and other dairy products; canned, frozen, and dehydrated fruits and vegetables; fruit and vegetable juices; preserves; meat; poultry; fish; soft drinks; beer and ale; wine; coffee; tea; spices and condiments; roasted nuts and nut butters; oleomargarine; shortenings and edible oils; and mayonnaise and salad dressings. [F.K.L.]

Animal feed manufacture. Feed processing involves a series of operations requiring considerable equipment, applied skill, and organized effort to procure, grind, and blend the variety of raw materials used in present-day feeds. Each grain is cleaned to remove foreign material, ground fine in a hammermill, and stored in bins equipped with feeders. Approximately 75% of the weight of most finished feed consists of processed cereal grains such as corn, sorghum grain, oats, or barley. Bran and shorts (by-products of the milling industry), distillers' dried grains (by-products of the fermentation industry), and soybean oil meal, cottonseed oil meal, or linseed oil meal (by-products of the oil extraction industries) are common constituents of formula feeds. The by-products a

other ingredients are also stored in bins. The remaining 25% of the feed consists of ingredients adding to the basic constituents special nutritional properties such as extra protein or fat content from either vegetable or animal sources, the essential amino acids, vitamins, disease-preventing drugs, hormones, and so on. *See* NUTRITION.

There are two systems of manufacturing formula feeds: the continuous and the batch methods. In the continuous method, the proportion of each ingredient required in the formula is compounded by percentage feeders continuously metering each required ingredient into a conveyor feeding into a mixer (Fig. 1). The batch system accumulates the feed in a batch scale hopper. Increments by weight of each ingredient are added to the batch which is discharged to a mixer that distributes the constituents uniformly throughout the mix. The finished formula feed may be packed as it comes from the mixer, or it may have liquids added, such as molasses, fish solubles, and stabilized animal fats, or it may be pelleted with or without added liquids. Pelleting compacts and binds all ingredients into small round or rectangular shapes from ⅛–1 in. in cross section. Pellets are considered more palatable, and assure the consumption of all constituents of the formula, reducing losses due to wind, dust, or spilling. When pellets are too large or compact, they are crushed with a pair of specially corrugated cracking rolls that break each pellet into two or three parts, called granules or crumbles. *See* ANIMAL-FEED COMPOSITION.

The finished feed, if necessary, may be dried to assure practical, safe storage without development of bacterial damage, or it may be packed in paper or burlap sacks of various sizes. Feed is also handled in bulk for delivery to feed lots in quantities that vary from 1 to 50 tons. Bulk shipments can be made by either truck or rail. [J.A.SH]

Pet foods. These are usually nutritionally complete products which adequately support growth and life even if fed exclusively.

Pet foods are made from proper mixtures of meat, fish, fowl, bone, cereal, milk, minerals, and vitamins. The last two named ingredients are added in quantities necessary to augment the natural foodstuffs. They may be prepared in various finished forms, canned, frozen, or dried.

The canned products are sterilized at high temperatures in order to permit storage without refrigeration. The frozen variety often is limited to raw meat or sea food. Dry preparations are stable at room temperature, though generally for not as long as canned items.

Those pet foods which include meat contain a fairly large proportion of highly nutritious organ meats as well as other meat products.

The manufacture of the canned variety is fairly standardized in that the grinding, blending, filling, and sterilizing procedures are well known. The dry variety, particularly dry dog food, can be manufactured in a number of different ways ranging from pelletizing of dry ingredients to hot-air and vacuum drying of moistened, expanded pastes and extrusions. Little or no preparation is needed for the frozen products since these are generally raw.

Aseptic canning of certain pet foods appears entirely possible. In this procedure the product is sterilized outside the can, then filled into a sterilized container, whereupon it requires no further cook-

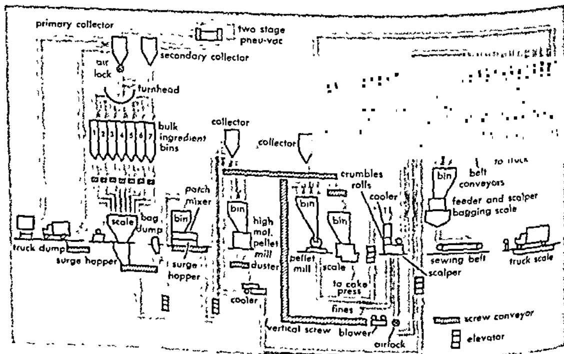


Fig. 1. Flow diagram of formula feed plant. (Sprout Waldron and Company, Inc.)

ing and is stable at room temperature for indefinite periods. [C.K.WI.]

Vegetable products. Vegetables are edible annual plants of which the immature succulent roots, bulbs, stems, blossoms, leaves, seeds, or fruits are eaten. Certain perennial nonwoody plants of which roots, stems, leaf stalks, or leaves are consumed are also classed as vegetables. Fruit vegetables such as pumpkin and tomato are technically fruits which are eaten as vegetables. Vegetables for processing are generally referred to as truck crops.

Canning. Successful canning of vegetable products ensures the destruction of food-spoilage microorganisms and enzymes with moist heat, and the storage in functional containers to prevent reinfection and chemical deteriorations. Both tin and glass containers are satisfactory and some plastic films may be used. The food spoilage organisms associated with vegetables are generally anaerobic spore-forming bacteria able to withstand several hours at boiling water temperatures. Vegetable products are therefore sterilized at temperatures between 115.6 and 121.1°C. In this temperature range, food qualities are successfully retained, and sterilization of 1-lb containers of vegetable products requires less than 1 hour.

Because most vegetables have pH values ranging between 4.5 and 7.0, there is danger from growth of *Clostridium botulinum*, and heat processes must be such that this organism is destroyed. It is possible to lower the pH value of vegetables to below 4.5, and thereby employ less-severe heat treatments. *See* BOTULISM.

In practice, vegetables for processing are harvested at their peak quality and brought to the canning factory. They are washed, sorted, prepared by such processes as peeling, trimming, cutting, or dicing, packed into containers which are usually filled with a hot dilute salt brine, sealed, given the necessary heat sterilization, cooled, labeled, and stored for subsequent distribution.

Freezing. Because of their high moisture content, vegetables usually freeze between −3.9 and 0°C. The more rapidly vegetables are frozen, the better is the retention of fresh qualities. Storage should be at −17.8°C, or lower. The successful freezing process involves a step for inactivation of enzymes in the tissues. The most satisfactory method to date has been by blanching, that is, heating the food to temperatures above 79.4°C. In addition to inactivating enzymes, heating wilts vegetables, facilitates packaging, sets food color, and removes the earthy odors of vegetables. Packaging is necessary to prevent the losses of moisture by sublimation, which causes dehydration of the tissues. In such a condition, termed freezer-burn, the tissues are irreversibly denatured.

Dehydration. For best storage stability, vegetable products must be dried to below 2% moisture content. Dehydration may be accomplished by tunnel, kiln, cabinet dryers, and vacuum dryers.

Vegetables to be dried are washed, sorted, trimmed, sliced, diced, and prepared for drying. Enzyme inactivation is required, and may be accomplished by blanching in hot water or steam. Prepared vegetable tissue may be given a treatment in dilute sulfite solutions, which protect against scorching damage during dehydration. Sulfites are removed during rehydration and cooking.

Dried vegetables may be compressed in volume and packaged, an important consideration in conserving storage space. Suitable containers are needed to protect the dried tissues from acquiring moisture from the environment, which accelerates deterioration.

Pickling. The fermentation of vegetables involves the conversion of sugars to acid with the aid of selected microorganisms. The addition of salt to the fermentation vat controls and establishes a selection of the organisms capable of growth. The fermented vegetable food products have clean, characteristic flavors and qualities. In combination with canning, fermentation of food products is adequate for preservation. *See* FOOD MICROBIOLOGY.

Juices. Vegetables are washed, sorted, trimmed, and pulped. During pulping, stems, seeds, and skins which may be present are removed. The pulp is further screened to yield a homogenous juice. The tomato is the most widely accepted vegetable juice, and is made by the extraction of juice and substantial pulp from red ripe tomatoes. The juice may have salt added as an optional ingredient. The juices of several vegetables, for example, tomato, carrot, parsley, and onion, may be combined into a vegetable cocktail juice. Canning is the most widely practiced method of preserving vegetable juices, although some juices have been dried successfully.

Purees. The pulped juice from vegetable tissues may be concentrated into puree. Such puree finds use in remanufacture into other food products, for example, tomato puree in the manufacture of bean sauces and in home cookery.

Pickles and relishes. Pickled vegetable products are made from fermented vegetable tissues. The fermentation is controlled as explained above. The fermented stock is washed free of excessive salt and packaged into containers in a spiced, flavored vinegar solution. The packaged pickles are then preserved by canning.

Pickled vegetables may be remanufactured into various forms of relishes, particularly where the shape and form of the fermented product has been damaged. Spiced, flavored vinegar solutions are added to combinations of pickled vegetables, either in large pieces or in coarse grind. Such relishes are preserved by canning, usually in glass containers.

Ketchup. Ketchup, or catsup, is a popular condiment, made by concentrating the juice and pulp of tomatoes, and adding salt, sugar, vinegar, and seasoning. Pulped tomatoes are concentrated by boiling or in a vacuum concentration unit to the desired consistency and solids content. The condiment is made into several grades, depending upon

the concentration made of the tomato solids. Fancy ketchup has more than 32% of total solids present, while standard grade has as little as 25%. The prepared condiment is preserved by canning, usually in glass containers for consumer uses. Chili sauce is a similar product to which the seeds have been returned and in which larger pieces of tomatoes are permitted to remain.

Fruit products. Fruits are the ripened ovaries of flowers, the edible portion being the fleshy covering over the seeds. The term fruit generally refers to tree fruits and berries. Being acid, fruits lend themselves to conventional methods of preservation and manufacture.

Canning. Because the main spoilage agents in fruits are non-heat-resistant microorganisms, canning preservation is functional. Exposure of prepared fruit products to temperatures near 93.3°C is generally adequate for their sterilization. The process involves the heating of manufactured fruit products in sealed containers of glass, tin, or plastic, in boiling water or steam, at atmospheric pressure. One-pound containers of fruit products usually are satisfactorily sterilized by a 20-minute treatment.

Freezing. The generally delicate nature of fruits requires rapid freezing techniques for best results, followed by storage of frozen products at −17.8°C or below. Sugar is usually a part of the packing medium, either mixed in dry or added as a syrup. Vitamin C (ascorbic acid) may be added to avoid color changes due to oxidation in ruptured tissues exposed to air. *See* ASCORBIC ACID.

Dehydration. Sun drying is a very widely used method of preserving fruits. Dehydration, or artificial drying, of fruits yields greatly improved products but is more expensive. Successful dehydration involves washing, sorting, trimming, slicing, or dicing fruit, exposure to burning sulfur fumes, or dipping into sulfite solutions, followed by the removal of the water in the tissues with dry hot air. The sulfur treatment is useful in protecting the delicate nature of fruit flavors and preventing discoloration due to oxidation. Dried fruits with moisture contents of less than 20% will usually keep for several years if protected from the elements and insects. Dried fruits with more than 26% moisture will mold and become inedible if not greatly sulfured. Drying may be accomplished by exposing the prepared, sulfured fruit to the sun or to drying in tunnels, in cabinet dehydrators, in kiln dryers, or in vacuum dryers.

Pickling. Fruits may be pickled by boiling in water until tender and placing in a spiced sugar syrup. The fruits are packaged into glass or tinned containers, filled with hot syrup, and sealed. If the fruit and syrup are above 82.2°C when packaged, no further processing is required. Filled containers should be inverted to sterilize the lids.

Jams, jellies, and preserves. The manufacture of fruit into jellies and preserves is one of the oldest and most important means of using large amounts of sound fruits which may not be well shaped or which for other reasons are unsuited for fresh market outlets.

Jellies are prepared by boiling the fruit, extracting the juice, filtering the juice, adding sugar, concentrating, packaging, and sterilizing.

The requirements of a fruit jelly are that the proper combination of acid, pectin, and sugar in water be reached so that a gel will form. This will occur when the pectin present amounts to 1%, the pH value of the substrate is approximately 3.1, and the concentration of sugar is near 65%. Too high a pectin content results in a very rigid gel, too low a content in a weak gel. If the pH value is below 2.9, a very weak gel forms, and no gel formation occurs above 3.5. Gel formation occurs with more than 60% soluble solids. Too high a sugar content results in a sticky gel. Acid, pectin, and sugar concentrations are interdependent; each must be in balance with the other variables to have acceptable gel formation.

Jams, jellies, preserves, marmalades, and fruit butters are products prepared from fruit juices or fruit with added sugar or both. After concentration by evaporation to a point where microbial spoilage cannot occur, the prepared products can be safely stored. Commercially, the prepared fruit products are filled into containers, usually glass, allowing the mass to set in the container.

A jelly is a semisolid food made from not less than 45 parts of fruit juice by weight to each 55 parts by weight of sugar. This substrate is concentrated to not less than 65% soluble solids. Pectin and acid may be added to overcome deficiencies that might occur in the fresh fruit. A jam has a requirement similar to jelly, with the exception that it is the fruit ingredient that is used rather than the fruit juice. Preserves retain the form of the original fruit and should consist of whole or cut fruit in a clear syrup of high sugar concentration. Otherwise the requirements are similar to those for jams and jellies. Marmalade is a product usually made from citrus fruit and is the jellylike product made from properly prepared juice and peel with sugar. It has the same requirements to achieve a gel structure as jelly. Fruit butter is a smooth, semisolid food prepared from a mixture containing not less than five parts by weight of fruit ingredient to each two parts of sugar, concentrated to not less than 43% soluble solids.

Fruits vary in their jelly-making ability. Crab apples, sour apple varieties which are not overly mature, sour berries, citrus fruits, grapes, sour cherries, and cranberries contain sufficient pectin and acid to yield good jelly naturally. Sweet cherries, quinces, and melons are rich in pectin but low in acid. Strawberries and apricots contain sufficient acid but are low in pectin. Peaches, figs, and pears are low generally in both acid and pectin. Commercially available pectin and edible acids permit a manufacturer to correct these deficiencies in fruits.

Fruit juices. Juices are made from fruit usually lacking good market quality, but of sound charac-

ter otherwise. Fruits are washed, sorted, trimmed, pulped, and extracted. The *juice collected is* screened and may be filtered. Some fruit juices are improved in appearance if clarified. This does not apply to juices of citrus, apricot, pineapple, and others which are more popular when cloudy. Apple, cranberry, cherry, and grape juice are examples of clarified juices which have good acceptance. They may be preserved by canning, freezing, dehydration, or chemical additives. Sodium benzoate is a preservative permitted by government regulation in fruit juices to the amount of 0.1%. Sorbic acid is another permitted preservative.

Grape juice is prepared from mature grapes which are washed, crushed, and pressed. The juice is pasteurized by heating to above 79.4°C, then stored to permit separation of suspended solids and tartrates, at temperatures below 0°C. The settled juice is separated from the sediment, then filtered, bottled, and sterilized.

Apple juice is made from washed apples which are crushed and pressed. The juice is filtered and bottled, with and without clarification.

Orange juice is extracted, and the juice is separated from the oil that may be released from the skin. The juice may be screened and canned, or concentrated in a vacuum system to 60°Brix. The concentrate may then be diluted with fresh juice to 42°Brix, packaged into cans, and frozen rapidly. The frozen concentrated juice is then reconstituted by adding three volumes of clean water.

Fruit purees. Fruit pulp may be concentrated and preserved as such by canning or freezing. Fruit purees are used mainly for manufacturing in the bakery and ice cream industries, or to reconstitute into fresh fruit beverages and fruit-juice-flavored drinks

Glacéd fruit. Fruit may be preserved by allowing its slow impregnation with syrup of increasing sugar concentration, developing such osmotic pressures within the tissues that microbial attack is prevented. This candied fruit is then washed, dried, and packaged. Candied fruit may be given a dip into syrup and again dried to develop a thin glaze of sugar coating; this product is called glacéd fruit.

[N.W.D.]

Confections. The per capita annual consumption of confections is 18 lb. The confectionery industry manufactures annually over 2,000,000,000 lb of carbohydrates with a value of more than $1,000,000,000. There are more than 2000 varities of confections and most of these have hundreds of different variations For example, there are as many as 500 different formulas for nougats.

About 70 farm products of varying carbohydrate contents furnish more than 90% of all the raw materials for the over 2000 varieties of confections. Common sweeteners used in confections are refined crystalline cane or beet sugar, brown sugar, liquid sugar, corn syrup, dextrose, sorbitol, starch, molasses, honey, and maple sugar. More emphasis is being placed on the addition of vegetable and animal proteins in confections, such as soya proteins, zein, yeast, gelatin, and pulverized dried beef. Also available are improved vegetable and animal fats. Fats and oil are important in fudge and nougat types of confections.

Usage factors. The major factors considered by a candy technologist in determining sweetener usage are differences in physical and chemical properties of various sweeteners, their relative prices, in some cases restrictions imposed by Federal or state regulations and, to a lesser extent, advertising and sales programs, in-plant handling problems, consumer preference, and psychological factors.

The most important physical and chemical properties, considered in their order of relative importance to the candy industry are (1) relative sweetness; (2) solubility and crystallization characteristics; (3) density of liquid sweeteners and moisture content of solid sweeteners; (4) hygroscopicity; (5) flavor; (6) fermentation and preservative properties; and (7) molecular weight, osmotic pressure, and freezing point depression.

Proteins and fats also vary considerably with respect to the above properties Requirements also vary widely, according to the qualities desired in the particular type or class of confection. The large variety of products turned out by the confectionery industry require a remarkable flexibility from the carbohydrates which serve as major constituents of these products.

Confections are divided into three categories: hard candies or high-boiled sweets, chewy confections, and aerated confections. The second and third categories are further subdivided into two classes, where the sugar solution is supersaturated (grained) or unsaturated (nongrained). Candies that grain are of crystalline structure and include the fondant types, such as cream centers, crystallized creams, fudge, pulled-grained mints, rigid-grained marshmallows, and soft- and hard-type pan centers. The nongrained candy group consists of marshmallows, taffies, chewy candies, such as nougats, caramels, and molasses kisses, jellies, and gums. There are many intermediate or hybrid types of confections combining the characteristics of both the grained and nongrained candies. Sugar is the universal graining agent. The regulators are corn syrup, invert sugar, sorbitol, and others that retard or prevent sucrose crystallization.

Sugar, or sucrose, has the highest rate of solution, and the smaller the crystal size the more rapid the solution. It forms highly supersaturated solutions which withstand supercooling. Most important of the chemical reactions of sugar is its ability to hydrolyze to produce invert sugar. The color developed in sugar solutions is a function of the pH, and color cannot be formed unless inversion has first taken place. Invert sugar syrups are known for being hygroscopic and retarding the crystallization of concentrated sugar solutions. *See* SUCROSE; SUGAR.

Corn syrups are noted for retarding and controlling the crystallization of concentrated sugar so-

lutions. For practical purposes regular corn syrup, high-conversion corn syrups, and enzyme syrups have equal effects and can be used interchangeably on a solids basis to control sugar crystallization in such products as fondant, fudge, or other grained confections. Total solubility of the many sugars found in corn syrup and moisture present in the confection minimize the crystallization of the sucrose. Singly, each sugar increases the tendency to crystallization, but this factor of total solubility in the liquid phase overbalances any such tendency. Dextrins add viscosity, which increases the body of nongraining confections. *See* DEXTRIN.

Dextrose, or glucose, possesses the ability to change solubility characteristics and modify the relative sweetness of confections. Dextrose also tends to crystallize more slowly than sucrose and the solution at the same concentration is less viscous. Sorbitol, derived from dextrose, is a sugar alcohol, which seems to have plasticizing properties in confections. Besides having a narrow humectant range, it has an effect on the sugar crystal which results in keeping candies soft for extended periods of time. It appears to be gaining in favor as a softening agent for both grained and nongrained confections. *See* GLUCOSE.

Hard candies. Hard candies or high-boiled sweets are essentially a highly supersaturated, supercooled solution of sucrose containing 1% or less of moisture. This solution, which may be likened to glass, is usually prevented from crystallizing by the addition of invert (or of invert formed in the batch) or corn syrup, or both, depending upon the properties desired. They are usually compounded in the ratio of 70 sugar to 30 corn syrup for open-fire cooking, and 60 sugar to 40 corn syrup in vacuum-processed batches. If the hard candy is to be relatively slow dissolving, a larger proportion of corn syrup is used, as in a banana caramel; if a quicker-dissolving product is wanted, such as a clear mint, smaller amounts of corn syrup are used. Corn syrup controls sweetness and reduces friability of hard candies. These are susceptible to fracture from mechanical shock, because of internal stresses induced by unequal cooling. Syrups of high dextrin content and low dextrose and maltose fractions are used to control hygroscopic properties. Formulations and processing techniques in the manufacture of hard candy confections are designed to make a product, of low hygroscopicity and different degrees of sweetness, which is ungrained, dense, and brittle in texture.

Chewy confections. Chewy confections of the nongraining type, such as kisses, caramels, gums, and jellies, are formulated with sucrose, corn syrup, fat, and milk solids. The ratio of sugar solids to corn syrup solids, plus 12–15% of moisture, is such that the carbohydrates remain in solution. The dextrins impart body or chewy texture, and in caramels the milk solids contribute flavor and texture. Natural and artificial flavors are accentuated by the sugar, invert sugar, and corn syrup. Fats, which are usually of the vegetable type because of their excellent shelf-life properties, are added to impart body and lubricating qualities. Emulsifiers, such as lecithin, monoglycerides, sorbitan monostearate, and polyoxyethylene sorbitan monostearate, are added to make the product more palatable Standard 42 dextrose equivalent (DE) corn syrup has widest application, as it contributes the desirable chewy characteristic without danger of excessive hygroscopicity.

Gums and jellies may be subdivided into two classes on the basis of the gelling agent, those utilizing starch and those utilizing pectin. Starch gums employ about 10% thin boiling starch which is gelatinized during processing by the free water in a sucrose–corn syrup solution which is cooked with it. It is essential to use excess water at the beginning of cooking to ensure complete gelatinization—that is, about 1 gallon of water for every pound of starch in the batch. No excess water is necessary, however, in a new continuous method. Sucrose and corn syrup are present in approximately equal proportions and are concentrated to about 76% solids before being cast into a cornstarch molding medium. On cooling, a firm, resilient, transparent gel results. Crystallization is prevented because of the high solubility of mixed sugar present and the gelatinized starch. Although standard 42 DE corn syrup is most commonly used for gum work, the enzyme-converted 53 DE type has advantages due to its humectant properties and extra sweetness. Up to 25% of the total sweetener may also be in the form of refined corn sugar, which promotes sweetness and reduces viscosity of the batch, enabling faster moisture evaporation and prolonging shelf life because of its hygroscopic character. However, a high ratio of dextrose to sucrose may cause graining in the piece due to the relatively low solubility of the former.

Pectin and low-methoxyl pectin jellies are compounds of sucrose, corn syrup, pectin, citric acid, water, and a buffer salt, concentrated by heat to a solids content of about 75%, which is necessary for proper gel formation. A 50–50 mixture of sucrose and corn syrup is conventional practice. This ratio effectively inhibits graining and gives the minimum difficulty from excess moisture absorption on the surface. Because the pH of pectin jellies must be adjusted to between 3.45 and 3.55 for proper setting of the pectin gel, significant quantities of invert sugar are formed from the sucrose, giving rise to possible subsequent sweating problems. Accordingly, a corn syrup of relatively low dextrose equivalent is the logical selection, so as to [illegible] the [illegible] sec- [illegible] ore [illegible] beneficial in starch gum confections.

Aerated confections. Aerated confections are of two basic types, those made from unsaturated solutions, and those made from supersaturated solutions which give grained confections. Aerated, unsaturated solutions, such as frappés and marsh-

mallows, are not necessarily chewy. Both of these products are formed with sugar and corn syrup in percentages rarely in excess of 45–55, respectively, combined with water and an aerating agent, usually albumen or soybean protein for frappés and gelatin for marshmallows. A diffusion method for continuous production of foam candy results in longer shelf life because of smaller uniform air cells, while less gelatin or other whipping agent is required because of lessened fatigue in process. Because of the high proportion of corn syrup and the protection afforded by the protein colloids, the system remains unsaturated with respect to sucrose. Concentration of the total solids without incurring supersaturation to a level safe from microbiological spoilage is made possible with the corn syrup or invert fraction, which may be either standard 42 DE or special 52 DE type in the case of uncoated goods, or special 52 DE and 63 DE enzyme or invert sugar in the case of coated marshmallow products. Molasses, honey, or maple sugar are used in all types of confections, mainly as a flavor. Honey has very good humectant properties.

Grained, aerated confections are the supersaturated solutions that form grained confections, such as creams, fondants, nougats, fudges, and grained marshmallow items. Fondants and creams are prepared by concentrating a sucrose, corn syrup, and water solution to about 85% total solids, using a ratio of sucrose in excess of that of corn syrup so that precipitation will occur when the supersaturated solution is seeded or crystallization induced by mechanical means. This generally implies a proportion of 80–70 parts of sucrose to 20–30 parts of corn syrup. The mixture is usually boiled to 114.4–116.7°C. Agitation or seeding is carried on at reduced temperatures of about 110°F, so that the crystals so formed will be of impalpable size. Some air is incorporated by the mechanical action of the equipment used in making the fondant. In the case of creams, the fondant is used as a seeding medium and a small percentage of egg or soybean albumen in the form of mazetta is used to impart some air. The function of the corn syrup portion is to serve as a humectant that will keep the products soft and palatable, and to permit concentration of soluble solids in the liquid phase to a level of about 80%, which will prevent growth of microorganisms. All three types of corn syrups have applications, the special 52 DE and enzyme-converted 63 DE imparting effective humectant properties. Invert sugar in combination with corn syrup gives excellent results. Refined corn sugar may be used in limited amounts, but it is not generally recommended because of its tendency to grain off in coarser crystals. Sorbitol with corn syrup seems to give a more uniform and whiter cream and fondant.

Fudges are similar to caramels, containing sugar, corn syrup, milk solids, and vegetable fats, but are slightly aerated by mechanical agitation with the aid of egg or soybean protein frappés. Soybean proteins are gaining favor because they bring out the color of the cocoa powder or cocoa liquor used in chocolate fudges. The ratio of sugar to corn syrup is higher, enabling precipitation of sugar crystals to form short, crystallized texture. All three types of corn syrup have applications, as in creams. Although the syrups of higher dextrose equivalents are used in chocolate-coated bars, the Quartermaster ration chocolate-type covered disks have 10% sorbitol to ensure a shelf life of the product of at least 2 years.

Grained marshmallows are produced with sugar, corn syrup, and an aerating agent such as gelatin or albumen. The corn syrup, invert sugar, or sorbitol serves to retain softness in a supersaturated, aerated sucrose system. Refined corn sugar (dextrose) may be incorporated to 20–25% of the total sweetener because of its hygroscopicity and higher fluidity which permit faster beating to the desired specific gravity of the product. Standard 42 DE and special 52 DE syrups are preferred for this item. Enzyme-converted syrups or invert sugar, if used, should be used sparingly to avoid excessive softening due to absorbed moisture. What applies for grained marshmallows applies for grained nougats. In this confection, gelatin is seldom used and the aerating agent is usually egg albumen, soybean protein, or a 50–50 mixture. Vegetable fats are used to promote smoothness and palatability to the confection. Since a high percentage of nougat confections is coated with chocolate or chocolate-type coatings, the 63 DE enzyme-converted corn syrups or invert sugar is used. [J.J.A.]

Bakery products. Bakery products can be made because of a unique property of wheat flour. Glutenin and gliadin, two of the major proteins of wheat, form a spongy mass called gluten when mixed with water. Gluten develops a gas-retaining structure in the doughs which causes them to increase in size when a leavening action is provided. This action occurs during a period of fermentation allowed after each batch is mixed and also during the first part of the actual baking of the product. Certain flavors developed as by-products of yeast fermentation increase the palatability of the baked product, and the dough is mellowed by the stretching action that is produced. If doughs have a high resistance to extension, considerable pressure will build up within the structure, causing a rupture of the cell walls. This creates irregularities such as holes or streaks, and may also reduce the volume of the baked product. The proper mellowing of doughs is necessary to reduce their resistance to expansion without creating a weakness to a point that causes any appreciable loss in gas-retaining properties. *See* FERMENTATION.

The most suitable flours are selected for each type of baked product, since they differ appreciably in protein percentage as well as in the quality of their gluten. Yeast-raised products are made with flours from spring and hard winter wheats. Cakes and certain biscuits which depend upon chemicals for the leavening action require the use of flours with an appreciably lower protein content, such as those obtained from soft wheats.

Bread production. Most bread is produced by the sponge dough process (Fig. 2). In this process, a portion of the ingredients is mixed with water to form the sponge as the first step in the operation. This mixture is allowed to ferment under controlled temperature conditions of some 80–84°F in a trough that may be 8–10 ft in length. The mixture is returned to the mixer and the remaining ingredients are added. The amount of mixing has a decided effect upon the finished product since it aids in dough mellowing, but excessive action can be too weakening in its effect.

The dough is then returned to troughs in the fermentation room for further fermentation. The trough containing the fermented dough is transferred mechanically to the divider, which divides the dough mass into pieces for eventual transfer to baking pans. Loaves are commonly 16–20 oz in weight and the dough piece needed to produce such a loaf must be approximately 18% heavier to take care of baking and cooling losses.

Since these dough pieces have cut and ragged surfaces, they are passed over a rounder that rolls them into smooth dough balls. The dough balls travel over a conveyor approximately 7–14 minutes through a controlled-atmosphere unit which is called the intermediate proofer. The time needed depends upon the characteristics of the flour in use. This relaxation period has been found desirable before proceeding with molding, the dough pieces into proper shape for transfer to pans. Molding begins by passing the dough between two or three pairs of rolls which gradually elongate the dough into a sheet. Each of these sheets is then curled, sealed, and deposited as a cylindrical mass into bread pans that pass underneath the discharge end of the molder unit at a controlled rate.

The pans are usually welded together in a set of four or five units. From the divider, these bread pans are transferred to the proof box. This is a chamber which is held at a temperature of 36.7–37.8°C with sufficient humidity to prevent any crusting on the surface. The fermentation is accelerated in the dough pieces by this temperature so that in some 60–70 minutes they have doubled in height and are ready for baking. This baking may be done on traveling flat bed ovens or on trays that travel vertically and horizontally in one or more cycles until the baking operation is complete. One-pound loaves usually require some 20–30 minutes.

After baking, the bread is dumped from the pans and is usually manually transferred to racks. These are then either placed in controlled cooling rooms or are left in the shop atmosphere until they reach about 35°C and are cool enough for slicing and wrapping. Slicing and wrapping machines are operated as a unit. Slicing is accomplished by band blades, so spaced as to give the desired slice thickness. The loaf is then wrapped in wax-impregnated paper which seals on the ends and bottom from the application of heat or in transparent wrappers that seal by a fusing action. Figure 2 pictures the order of these operations in mechanized bakeries.

Many smaller bakeries do not have all this equipment, and one or more of the operations listed are performed by hand. A straight dough process, involving only one mixing operation and fermentation period, is often used in such establishments. Variety breads containing variable proportions of whole wheat or rye flours are made in considerable quantity, with an appreciable production of loaves containing raisins.

Both hard and soft rolls are produced in a variety of shapes and sizes. There is also a wide range in the proportion of ingredients used for the production of these items, and one class of such products is commonly called sweet goods because of high sugar and fat content. Special equipment is often used for the make-up operations of the roll department, but the principles regarding flour types used and the fermentation needed are essentially the same as for white bread production.

A revolutionary method of bread production has been introduced in several bakeries. It does not make use of the mellowing effect of sponge or dough fermentation. A liquid ferment is set, containing the yeast and yeast nutrients in order to obtain an active fermentation and to develop flavor.

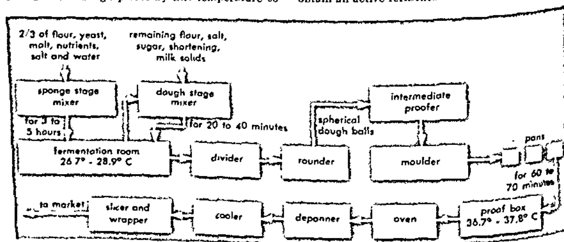


Fig. 2. Bread production flow chart.

This broth is then continuously proportioned into a machine along with the flour and other ingredients needed. This machine is a premixer or blender, and the blended dough is fed into a developer which subjects it to a very vigorous mixing action. This action is carried to such a point that the gluten structure, as formed in the premixer, loses most of its elastic properties. This means that the doughs, which are extruded directly into pans, have a minimum resistance to expansion and still have satisfactory gas-retaining properties. Bread made by this continuous mixing process usually has a more uniform cell structure than that made by conventional methods. This simplified method of bread production is gaining supporters, with a resultant increase of suppliers of such equipment.

Other products. Cake production is a simpler operation than bread making. Cakes differ widely in characteristics, but since they are made without yeast no fermentation periods are involved. Instead of doughs, batters are used for cake production, the basic difference between the two being the thinner consistency of batter due to the higher liquid content. Cakes may be divided into foam and shortening types. In the shortening type, the leavening action comes from the reaction of sodium bicarbonate and an organic acid. These may be added as separate ingredients or in combination as baking powder. In the foam-type cakes, a high proportion of egg white is used, which aids the batter in retaining the air incorporated during the mixing operation. The expansion of this air during the baking operation normally gives all the needed volume without the use of any chemical leavening. Handling methods are essentially the same for all types of cakes. The batter is mixed and deposited in pans by mechanical means, which go directly to the oven for the baking operation. Cakes are then transferred to racks for cooling and are wrapped on machines specially designed for that purpose.

This usual method of cake production is a batch operation as far as mixing is concerned. Other mixers are now also in use to which ingredients are continuously proportioned and which deliver an uninterrupted stream of batter into the cake pans.

Doughnuts are made in bakeries but are fried in fat instead of being baked. Some doughnuts make use of yeast and a fermentation reaction for lightness, while cake types are chemically leavened. Special machines are used for scaling off the desired weights, shaping the pieces, and depositing them into the frying fat.

Most cookies are chemically leavened and so are similar to cakes except for adjustments necessary because of differences in shape and size. Ordinary crackers are made from yeast-fermented doughs and with the use of relatively low-protein flours. An adjustment of the developed acidity by the addition of soda produces the type of cracker that is most common.

Pies make up a considerable proportion of the output of some bakers, especially retail ones. Pies may have only a bottom crust or also a second one over the filling. A large number of fruit and other fillings are used. Pie crusts can best be made from relatively low-protein flours and must contain a high fat percentage to have the desired tenderness.

Amounts produced. The accompanying table gives information on the size of the baking industry in terms of poundage produced. The source is the U.S. Bureau of the Census, 1954 compilation. Other surveys have indicated no appreciable change in production figures during recent years, so the data can be accepted as accurately reflecting current production of the baking industry, exclusive of retail, single-shop bakeries. The latter are not a major factor in the production of white bread and other items on the list. [O.S.]

Table 1. Annual baked goods production (not including retail single-shop bakeries)*

Product	Thousands of pounds
White pan bread	8,614,290
White hearth bread	252,820
Whole wheat and other dark wheat breads	996,029
Rye breads	620,278
Raisin and other specialty breads	327,975
Rolls—bread type	1,501,624
Sweet yeast goods	630,812
Doughnuts—cake type	375,929
Soft cakes	1,035,031
Cookies—handmade	62,288
Cookies—machine-made	1,139,111
Crackers	1,122,225
Pretzels	88,220
Pies	724,137

SOURCE: U.S. Bureau of the Census

Sauces. A sauce is an appetizing dressing for food and is used as a condiment, relish, or vehicle. Sauces may be used at the table or in cooking. When used at the table they serve as condiments; when used in cooking they may also serve to bind ingredients together.

Important sauces are made as follows: Tomato catsup is tomato pulp strained free of skins and seeds and concentrated, with vinegar, sugar, and spices. Worcestershire sauce is a combination of vinegar, hydrolized proteins, vegetables, and fruits which is allowed to age in barrels to develop flavor blend. Chili sauce is tomato pulp with seeds left in, concentrated and with added vinegar, sugar, and spices. Chop suey and soy sauces contain proteins such as wheat, corn, or soybean, hydrolized by carefully controlled acid hydrolysis, which may have added condiments, molasses, and caramel color. Pepper or hot sauce is made by concentrating chili or tabasco peppers and adding vinegar and salt. Meat or barbecue sauce is a combination of tomato pulp, soy base, vinegar, and spices. Mustard is made by grinding or milling mustard seeds, vinegar, and spices. Applesauce is a combination of cored, sliced apples which are cooked with sugar, water, and spices. Cranberry sauce is made from cranberries, sugar, and water. Special sauces such as hollandaise, bearnaise, tartar, and chutney are

manufactured by combining specific ingredients. Sometimes homogenization is necessary, depending on the final product desired

Most sauces are packed in glass containers with care given to the choice of closure.

Sauces are made by using specific types of mixing and homogenizing equipment. These may be vertical-type mixers, colloid mills (horizontal or vertical), pressure-type homogenizers, emulsifiers, or supersonic equipment. The purpose of these is to supply uniformity of final product by better emulsification and particle-size reduction. Homogenizing or milling is of importance in many of the above sauces because where there is the addition of a vegetable gum or stabilizer, proper dispersion must be secured These gums and stabilizers provide for smoother texture of the sauce and diminish the possibility of separation into specific phases, thereby improving taste and appearance of the sauce. [A.J.F.]

Soups. Canned soups are manufactured by conventional canning processes They may be clear, cream, thickened, vegetable, or vegetable-meat type. They are usually concentrates which are cooked in batches. Heat processing may be short or long. The hot batch is filled directly into cans, which are hermetically sealed and sterilized.

Frozen soups are cooked concentrates prepared batchwise as for sterilized soups After filling, the cans are chilled and quick-frozen Storage is at 0°F or lower until distributed Fish base soups, in particular, are well-suited as frozen varieties, since conventional sterilization would greatly impair their flavor.

Dehydrated soups are mixtures of dry ingredients with dehydrated stock and either dehydrated vegetables or meat. After careful blending, the mixes are filled into flexible-wall bags or paper boxes which are automatically sealed. Modern packaging developments have provided wall structures which satisfactorily protect dehydrated soups from moisture uptake. [J.H.N.]

Poultry products. Over 5,000,000,000 lb of poultry was delivered to the American consumer in 1957. A major portion was fresh (ice-packed) poultry either as whole chickens or cut-up parts. About one-third of this total was delivered to the consumer as frozen whole or cut-up poultry. Canned poultry, either whole or parts, and as ingredients in such items as chicken pies and TV dinners, constituted 227,000,000 lb.

Commercial poultry production in flocks of 25,-000–100,000 chickens accounts for most of the production. In modern poultry plants, birds are dispatched humanely and allowed to bleed thoroughly before going to the scald tank to facilitate removal of feathers. Feathers are removed by mechanical equipment with a minimum of hand work required. While suspended from moving sanitary conveyors the poultry is further processed under close inspection. All viscera is removed along with heads and feet. The eviscerated poultry is cooled to below 1.7°C in less than 2 hours in tanks containing large quantities of cracked ice. After cooling, the poultry is cut into parts and packed into boxes for freezing, packed in ice whole and ready to cook, or frozen for distribution. Poultry to be canned is processed in a similar manner, after which it is pressure cooked. The meat, removed from the bones, is placed in cans and sterilized in the usual manner or used as ingredients in chicken pot pies, turkey pot pies, and so on. Specialty items, such as roast stuffed turkey, precooked fried chicken, and precooked complete dinners are manufactured by many firms. [R.H FO.]

Meat products. The manufacture of meat products includes those processes which prepare the product for consumption and increase the stability, improve the texture, increase the convenience, or alter the flavor, color, and appearance of various meat items. Various processes are employed, depending upon the desired result. These include cooking, freezing, comminuting, curing, smoking, drying, canning, and packaging. Salt, sugar, sodium or potassium nitrate or nitrite, spices and flavorings, as well as various enzymatic agents and other additives often are used. These are important both to the packer and to the consumer. The packer benefits because of the increased utilization of meat cuts and the greater stability given perishable raw materials. The consumer benefits in having available a greater variety of items at a reasonable price and in a range of textures and flavors.

The manufactured meat products can be grouped as follows: cured and pickled, cured and smoked, tenderized fresh, frozen, and canned.

Cured and pickled meats. Cured meats are those items which have had combined with them salt, sodium or potassium nitrate, and sodium or potassium nitrite. Sugar and spices are optional ingredients. The cure may or may not be added as a water solution.

The salt functions as a preservative, while the nitrate and nitrite combine with the meat pigments to form fairly stable colored compounds.

Cured meats can further be divided into two major categories, those which have been heated or cooked and those which have not. An example of cured meat which has not been heated is corned beef brisket. This item is produced by the use of a liquid curing or pickling solution. The beef brisket may be pumped with the curing pickle by means of a single- or multiple-needle injection device and then be immersed in a pickling solution for a period of 1–7 days.

In the category of cured and smoked meat products is dried beef, which is cured by being immersed in pickling solution and then heated, dried, and smoked for a prolonged period of time.

Cured and smoked meats. In the category of cured, smoked, and cooked meats is a broad line of sausage products such as frankfurters and bologna, which differ from cured and pickled meats in that they are prepared from finely chopped or comminuted meat to which have been added salt, sugar, spices, and flavorings (Fig. 3). They may

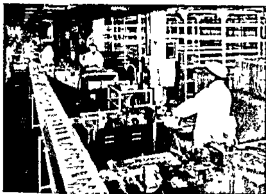


Fig 3 Frankfurter packaging line. In foreground, skinless franks are automatically peeled of the casings used in their processing and ejected onto moving belt that hastens them to weighing and packaging operations in rear. (*Armour and Co.*)

also include such items as cereals, milk powders, protein hydrolysates, and other substances.

These products are not pickled since cure penetration is obtained during an extensive mincing or chopping procedure. To secure the low temperatures often necessary for stability during the heating process, suitable amounts of ice are incorporated during the chopping operation. Ice also introduces moisture and thus increases the acceptability of the end product by assuring proper juiciness.

The products are usually heated in the smoke house to approximate internal temperatures of 65.6–76.7°C. Hardwood smoke is introduced, and the products are smoked for a length of time sufficient to impart the characteristic smoked flavor. This smoking operation further increases stability of the end products by depositing on the surface a certain quantity of bacteriostatic agents.

Since these items have been heated to temperatures sufficiently high to inactivate many undesirable parasites, they are generally considered suitable to eat without any further heating.

Tenderized fresh meats. Enzyme tenderization of fresh meats, particularly of certain beef cuts, has been a practice of long standing. A proteolytic enzyme like papaya, in fairly purified form, is usually used. The meat products are dipped in enzyme solutions and then frozen if they are to be sold in interstate commerce.

Frozen meats. Many frozen meats fall in the category of processed meat products since the processes employed in their manufacture increase the utility or the convenience of these items. Examples of this procedure are the various breaded items such as veal cutlets, pork chops, meat sticks, and related items (Fig. 4).

The cooking and rapid freezing of certain meat dishes is a well-established practice and an example of manufacturing processes which increase the convenience of meat products for consumer use. In this category are such items as meat pies, precooked meat entrees, and various complete dinners. Generally the precooked meat items hold up better in frozen storage if they are covered with a gravy, since this covering of gravy apparently tends to reduce oxidation.

Canned meats. These are meats that are preserved by heat sterilization while enclosed in cans or glass jars. They may or may not be precooked prior to being placed in the container. If finely divided materials are to be processed and if it is desirable that they remain in this finely divided state, it is necessary that these items be precooked prior to being placed in the container. Such precooking usually is only sufficient to denature the proteins so that they do not exhibit any binding properties.

The containers are processed at 104.4–126.7°C for periods of time sufficient to inactivate most bacteria and to cause the heat-resistant type of bacteria to become dormant. This kind of treatment results in a commercially sterile product. The shelf life of these items is theoretically unlimited, although in practice it seldom exceeds 2 or 3 years.

It is often desirable to use meats which are slightly tough, since high-temperature pressure cooking causes tender meats to fall apart.

Certain canned meats are produced without inducing complete sterility. These include items such as large canned hams, luncheon meat, and spiced ham. These items generally are reacted with curing ingredients and then processed to temperatures of approximately 71.1°C while in the can. Such items generally require storage at refrigeration temperatures of 4.4–10°C and are very stable at these temperatures, having shelf life of approximately 2–3 years or longer.

Fats. These are important by-products of slaughtered livestock and the manufacture of meat products. As raw materials, fats form the base of a great variety of finished fat and oil products.

Heat treatment or rendering separates the fat from the fatty tissue on which it is found in the [illegible]

[illegible] use include low-temperature dry rendering, high-temperature dry rendering, and steam rendering.

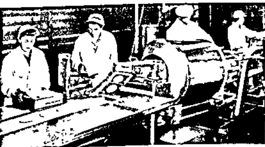


Fig. 4. Stainless steel breading machine dips frozen veal cutlets in batter, right background, then covers them with breading as they pass through rotating drum. (*Armour and Co.*)

Table 2. World catch of seafood by countries and by species groups, 1956*

Country	Million pounds (round weight)	Per cent
Japan	10,500	16.2
United States and Alaska	6,472	10.0
China (mainland)	5,820	9 0
U.S.S R.	5,769	8.9
Norway	4,693	7.3
Canada	2,374	3.7
United Kingdom	2,316	3.6
India	2,232	3.5
Germany, Federal Republic	1,699	2.6
Spain	1,651	2.6
Indonesia (1955)	1,436	2.2
France	1,186	1.8
Iceland	1,140	1.8
Portugal	1,039	1.6
Denmark	1,021	1.6
Angola	927	1.4
Philippines	917	1 4
North Korea	814	1.3
South Korea	752	1.2
Netherlands	657	1.0
Union of South Africa	633	1.0
Pakistan	611	0 9
Southwest Africa	591	0.9
Peru	551	0.9
Italy	482	0 7
Thailand	480	0.7
Sweden	434	0.7
China (Taiwan)	426	0.7
Chile	415	0.6
Other	6,593	10.2
Total	64,661	100 0

Species groups	Million pounds (round weight)	Per cent
Herring, sardine, anchovy	15,410	23.8
Cod, hake, haddock	10,758	16.6
Fresh-water fish	6,768	10.5
Tuna, bonito, mackerel	3,770	5.8
Salmon, trout, smelt	1,521	2 4
Flounder, halibut, sole	1,411	2.2
Miscellaneous marine teleosteans	6,019	9.3
Mixed and unidentified fish	11,552	17.9
Mollusks	4,079	6 3
Crustaceans	1,720	2.7
Shark, ray	617	0.9
Aquatic animals, N.E.S.	154	0.2
Aquatic plants	882	1.4
Total	64,661	100 0

* Includes all aquatic mammals (except baleen and sperm whales) and the commercial and subsistence catch of teleostean species (ganoids), elasmobranchs, crustaceans, mollusks, and other aquatic animals and plants. From FAO, *Yearbook of Fishery Statistics, 1955 1956*, vol 6, 1957, and E. A. Power, *Fisheries of the United States and Alaska, 1957—A Preliminary Review*, U.S. Fish Wildlife Serv., Fishery Leaflet 393, rev., 1958

In the manufacture of lard a number of processes follow rendering. These include decolorizing or bleaching, deodorization, chilling, plasticizing, and packaging. In some cases lard is hydrogenated. Recently the catalytic rearrangement of lard molecules has been widely adopted for the purpose of improving functional properties.

Rendered beef or mutton fats are termed tallows or oleo stocks, depending upon the rendering method. Tallow is processed similarly to lard Oleo stock is grained and pressed to yield oleo oil and stearine. Rendered lard is sometimes grained and pressed, but this procedure is diminishing

Refined lard enjoys great demand as a shortening and, with edible tallow, is increasing in use as an ingredient in the formulation of compound shortening products. Both lard and oleo products are used in the manufacture of margarine. See FAT AND OIL, EDIBLE. [C.K.W.]

Fish and seafood products. A great variety of food is prepared from many species of fish (Pisces) and shellfish (Mollusca and Crustacea). These are preserved by refrigerating, freezing, canning, salting, smoking, pickling, dehydrating, or combinations of these processes. In many countries they serve as a principal source of protein and an important source of fat, minerals, and vitamins in the diet.

In the United States, in 1957, the value of seafood products to the processor was $667,267,000 and to the retailer, $1,091,000,000. There were 84,050 vessels and 4225 shore establishments in operation which employed directly 245,300, and indirectly 310,000 people. The per capita consumption was 10.2 lb.

Fresh seafood products. These are highly perishable. Deterioration proceeds in direct proportion to delay in initial chilling and to storage temperature (degrees above 0°C). Chilling or refrigeration of round and dressed fish usually is accomplished by surrounding the product with flaked, chipped, or shaved ice, but refrigerated brine systems have been introduced successfully in

Table 3. Estimated use of fish catch, 1956*

Use	World		United States and Alaska	
	Billion pounds (round weight)	Per cent	Million pounds (round weight)	Per cent
Marketed fresh	27.2	42 0	1,892 2	29 3
Frozen	4 9	7 6	594 0	9 2
Cured	15.7	24 3	78 1	1 2
Canned	5 7	8 8	1,495 6	23 2
Reduced to meal and oil†	9 0	13 9	2,341 2	36 2
Miscellaneous purposes	2 2	3 4	57 9	0 9
Total	64.7	100 0	6,459 0	100 0

* From FAO, *Yearbook of Fishery Statistics, 1955-1956*, vol 6, 1957.

† Includes only whole fish destined for manufacture of meal and oil

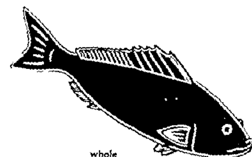
some fisheries. Fillets, steaks, and shellfish meats are packed into metal or glass containers, plastic bags, or cardboard cartons which are then surrounded by ice. Some market forms of fish are shown in Fig. 5. Extension of fresh storage life by such chemical preservatives as benzoates, nitrites, and organic chlorine compounds, and by antibiotics like chlortetracycline and oxytetracycline incorporated into ice has been proposed, but their use has not been widely accepted. Irradiation pasteurization followed by refrigerated storage appears to be successful for this purpose and is being further investigated. *See* CHLORTETRACYCLINE; OXYTETRACYCLINE.

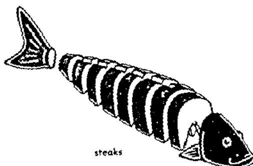
In most countries, strictly enforced regulations control public health problems involved in production and marketing of shellfish. Oysters and clams are marketed in the shell or as fresh meats. Only the adductor muscle of scallops is marketed. Crabs are cooked and sold as whole, eviscerated, and picked meat. The meat is washed in dilute brine to remove shell fragments and may be pasteurized to reduce bacteria count and inactivate enzymes. Most lobsters are marketed alive, although some are sold as chilled boiled lobsters or lobster tails. Prawns usually are beheaded at sea since only the tail portions are edible. Pacific Coast shrimp are shelled by mechanical peelers or by hand, and sold as cooked meats.

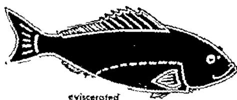
Frozen fish and shellfish. The production of these items is increasing throughout the world. In the United States, fillets, blocks, and steaks are of greatest importance.

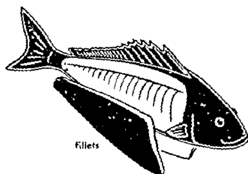
Whole and packaged seafoods are frozen in sharp freezers with and without air circulation, in plate freezers under pressure, and in air-blast freezers. Freezing at sea by direct immersion of whole fish in refrigerated brine has long been used in the tuna and salmon fisheries. Filleting and freezing groundfish aboard factory ships is becoming increasingly important in Europe.

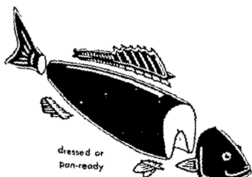
Adverse changes occurring in frozen fish include dehydration, discoloration due to oxidation and decomposition of blood pigments, rusting due to oxidation of fat, loss of flavor, development of off flavors, acquisition of undesirable odors, and











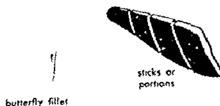


Fig. 5. Common market forms of fresh fish. (U.S. Fish and Wildlife Service)

denaturation of protein. Their development is inhibited by use of packaging materials with low moisture and oxygen transmission rates, packages without air spaces, vacuum and gas packaging, proper warehousing, control of temperature and humidity, and use of ice coatings or glazes.

For fish of good initial quality, maximum storage limits are 10 months for lean and 6 months for oily fish at the usual commercial storage temperature of −5–0°F. Storage life is decreased at temperatures above 0°F and greatly extended at lower temperatures.

Precooked and prepared frozen fishery products have become a major factor in the industry since 1953. While many products (portions, cakes, dinners, main dishes, and pot pies) are available, fish sticks and breaded shrimp are the principal production items (Fig. 6). The 1957 United States production of fish sticks was 52,400,000 lb, 89% as precooked and 11% as uncooked sticks. More uncooked breaded shrimp are produced than cooked items.

Canned seafood products. These are seafood [illegible]

The United States production of [illegible] is shown in Table 5.

The operations in canning seafoods are similar to those used for other protein foods. Specific details vary with different species and location. The

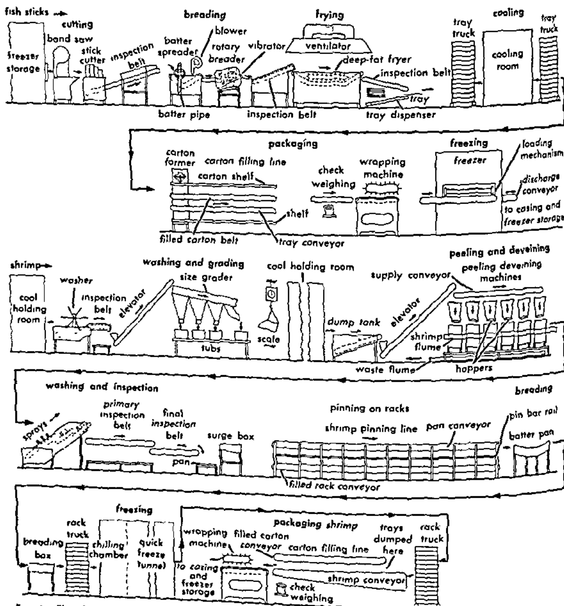


Fig. 6. Flowsheet of fish-stick and breaded-shrimp production lines. (Reprinted from Food Engineering, 27(2): 118–121, 1955)

Table 4. World production of canned seafoods, 1956[a]

Item	Million pounds
Pacific salmon[b]	341.0
Herring, sardine, anchovy[b]	919 6
Tuna, bonito, mackerel[b]	706 2
Miscellaneous fish products[c]	
U S S R[d]	514 8
Other countries	217.8
Crustaceans and mollusks[d]	167 2

[a] From FAO, *Yearbook of Fishery Statistics, 1955–1956*, vol 6, 1957.

[b] Excludes U S S R

[c] Not included elsewhere for countries other than U S S R

[d] *Includes all kinds of canned products prepared from fish, crustaceans, mollusks, and miscellaneous aquatic invertebrates*

Table 5. United States production, 1957, and record pack of certain canned seafoods*

Item	Pounds per case	1957 thousand standard cases	Record pack year	Thousand standard cases
Tuna†	‡	11,891	1957	11,891
Salmon	48	3,185	1936	8,965
Sardine				
Marine	20 3	2,266	1950	3,844
Pacific	45	498	1950	5,071
Mackerel and jack mackerel	45	1,327	1947	1,755
Anchovy	31.25	440	1953	1,062
Crabmeat	19 5	155	1948	221
Oysters	14	499	1937	760
Shrimp	15	602	1937	1,435
Clams				
Whole and minced	15	654	1957	654
chowder	30	1,218	1953	1,293
Squid	48	216		
Pet food and bait	48	7,130	1957	7,130
Other	Various	863		

* From E. A. Power, *Fisheries of the United States and Alaska, 1957—A Preliminary Review*, U S. Fish Wildlife Serv, Fishery Leaflet 393, rev., 1958.

† Includes production of tuna canned in American Samoa, Hawaii, and Puerto Rico.

‡ Solid, 21 lb, chunk, 19 5 lb; and flake and grated, 18 lb.

condition of the raw material (freshness, size, sexual maturity, and substances upon which the animals have been feeding) and the workmanship affect the quality of the canned item. Proper cooling is essential to avoid overprocessing or stack burn.

Discoloration of seafoods and containers is a serious problem. Sulfur, released as hydrogen sulfide, combines with the iron of the can forming black iron sulfide. This may be prevented by the use of parchment liners, increasing the acidity of the product, or enameling the interior of cans with materials containing zinc. Zinc reacts with the hydrogen sulfide to form white zinc sulfide. The most common defect in crab is blueing, caused by the reaction of oxidized copper in the hemocyanin in the crab's blood with ammonia in the flesh resulting in a copper-ammonium complex. Live slaughtering and bleeding of crabs minimizes this defect, while dipping crabmeat in brine containing zinc or aluminum salts inhibits this discoloration. The development of discoloration in canned tuna is caused by abnormal reactions of heme-containing compounds.

Occasionally, glasslike crystals of magnesium ammonium phosphate hexahydrate (struvite) are formed in canned seafoods. Their development may be inhibited by the addition of sodium hexametaphosphate.

Cured products. These include seafoods which are processed by drying, salting, smoking, pickling, ripening, and combinations of these. Curing is the most important means of seafood preservation other than chilling in every country except the United States and Canada. Preservation is effected by reduction of moisture, increase in acidity, or by the development of unfavorable temperatures. Throughout the world more herring is cured than any other species.

Drying of seafoods is done by natural or mechanical means. Natural air drying attempts to retain the form of the product. Fish are split, cleaned, boned, and hung on racks to dry. Shrimp, the most important product dried in the United States, are cooked in brine, peeled, and then dried on screens. Dehydrated fish are those which are cooked, flaked, and ground prior to drying by mechanical means. Freeze-drying has been investigated and found suitable for use with seafoods. Drying by sublimation from the frozen state using microwaves as an energy source is a recent experimental innovation.

Salting, the most important method of curing, preserves by extracting water (osmosis) from the product. Purity of salt and maintenance of low temperatures during curing are essential. Dry-salting procedures are used for lean fish such as cod. Cleaned and split fish are piled in alternate layers with salt, and the brine which forms is permitted to drain away (kenching). Fat fish are brine-salted to minimize contact of fish with air. Mild-cured salmon is the most expensive item produced by brine salting.

Spoilage occurs in salted fish when halophilic bacteria cause reddening of the surface of the fish or molds cause dun (chocolate-brown spots) or other colored colonies on the surface. Most salted products must be held under refrigeration. *See* BRINE, MICROBIOLOGY OF.

Smoked seafoods are preserved by salting and drying, although compounds (phenols and formaldehyde) which inhibit spoilage are deposited on the fish from the smoke. Hardwoods are used exclusively. Liquid smoke and salt impregnated with wood-smoke chemicals are sometimes used as flavoring agents. Lightly smoked seafoods, frequently artificially colored, must be maintained under refrigeration or further preserved by canning or freezing. Hard-smoked or dry products are comparatively stable at room temperature. Most spoilage is due to mold growth on surfaces.

Seafoods are cold-smoked (below 32.2°C) and hot-smoked or barbecued (cooked at smoke hous

temperatures of 65.6–121.1°C). Herring is the only smoked seafood that is not a luxury. Hard-smoked or red herring are heavily salted and cold-smoked until hard and dry. Bloaters are round herring, lightly salted and cold-smoked for a short time. Kippers are split and eviscerated fish, lightly salted and cold-smoked for a short time. Buckling are large, fat, round herring, lightly salted, cold-smoked for a short time and then cooked by hot smoking.

Pickled seafood products are preserved by organic acids either produced by fermentation of the raw material or by addition of vinegar (acetic acid). Added spices exert a slight preservative action. Pickled seafood products are the most important source of animal protein in many Asian countries, of major importance in Europe, but minor in the United States.

Fish meal and oil. These are produced from the inedible portions of fish, though some industrial or scrap fish, menhaden in particular, are caught only for reduction purposes. Table 6 shows the production of fish meal and oil in the United States.

Fish meal is produced by a dry-rendering or batch process and by a wet reduction or continuous operation. Solvent extraction methods are also used. In the batch operation, the ground raw material is dried under vacuum in a steam-jacketed drier equipped with rotating paddles. The dried meal is pressed hydraulically to remove excess oil which is then purified by use of settling tanks or centrifuges.

The raw material for wet-rendering is ground, cooked by steam injection, and pressed in continuous screw-type equipment to remove most of the water and oil. The meal is dried in direct flame, steam-jacketed, or steam-tube driers. The press liquor is centrifuged to separate the oil from suspended and dissolved protein materials (stickwater), which may be concentrated to 50% solids and sold as condensed solubles or put back in the meal during drying.

Table 6. 1957 production of fish meal, oil, and solubles in the United States*

Item	Unit	1957	Record	
			Year	Quantity
Meal				
Menhaden	Tons	172,388	1956	210,582
Sardine, Pacific	Tons	1,474	1936	121,739
Herring, Alaska	Tons	7,824	1937	18,816
Herring, Maine	Tons	4 956	1952	6,665
Tuna and mackerel	Tons	25,716	1956	26,266
Other	Tons	50,103		
Oil				
Menhaden	Gallons	15,797,919	1956	22,428,082
Sardine, Pacific	Gallons	87,495	1936	26,131,439
Herring	Gallons	1,697,447	1937	5 623,045
Tuna and mackerel	Gallons	738,279	1950	884 434
Other	Gallons	1,823,111		
Homogenized-condensed fish	Pounds	56 786 000	1956	59,576,000
Fish solubles	Pounds	187,760 362	1956	198 081,479

* From E. A. Power, *Fisheries of the United States and Alaska, 1957—A Preliminary Review*, U.S. Fish Wildlife Serv., Fishery Leaflet 393 rev., 1958

Fish meals contain up to 70% protein, less than 10% moisture, and less than 15% oil. They are used as supplements in animal feeds as a source of high-quality protein, B vitamins, the animal protein factor, and minerals. High-quality fish meals are used for human consumption in Africa and Asia.

Homogenized condensed fish is a concentrated partial enzymic acid or alkaline hydrolysate produced from fish waste or scrap fish which is used as an animal feed supplement in place of fish meal.

Fish body oils produced during the manufacture of fish meal are highly unsaturated. Industrial quality oils are used in paints, inks, and for ore flotation. Edible-quality oils are used in margarine and for addition to certain canned seafood items

Fish liver oils are the principal natural source of vitamin A and contain considerable quantities of vitamin D. They are used in animal feeds, in pharmaceutical preparations, and for fortifying foods such as margarine. Their production has decreased since production of synthetic vitamin A in 1949. *See* VITAMIN A; VITAMIN D. [J.A ST.]

Dessert manufacture. This is the commercial preparation of sweet foods to be served at the end of a meal.

Canned fruits such as peaches, cherries, and pineapple are packed either in a sugar syrup or in a sugarless water solution with or without synthetic sweeteners, and with or without thickeners to simulate a sugar syrup. The latter variety is due to the emphasis on low-calorie and dietetic foods.

Many convenience desserts, such as rice pudding, Indian pudding, and crepes suzettes, are canned for home use. Pie fillings are packed in cans and packages. Frozen desserts include fruits like strawberries and rhubarb, specialty items like pies and puddings, and many types of ice cream and related desserts. Among dehydrated desserts are found dried fruits (figs, apricots, plums). There are gelatin desserts made by combining sugar, plain gelatin, citric acid, sodium citrate, flavor, and color. Pudding powders, such as chocolate pudding, may be regular (requiring cooking) or instant (requiring mixing only). These contain powdered sugar (dextrose), starches (pregelatinized in instant pudding), flavor, and color.

Condiments. These are agents used to enhance the flavor of a food, to alter or supply a distinctive flavor when none exists, and to change the appearance of foods by giving color (like paprika, or curry powder). Condiments may include many varying types of materials such as (1) spices like pepper, mustard, ginger, nutmeg; (2) herbs like basil, marjoram, tarragon, parsley; (3) seeds like poppy, anise, dill; (4) chemicals like sodium chloride and monosodium glutamate; and (5) sauces made from cooking wines, vinegars, and relishes which are also considered to be condiments.

Properties of most condiments are distinct because of their volatile characteristics or the substances of definite chemical composition therein

Standards have been promulgated by the Federal Food and Drug Administration for the more

important spices, herbs, and seeds as listed above. These consist of specifications for percentages of crude fiber, total ash, volatile oils, starches, and other basic constituents of each specific item. The purpose of such standards is for the maintenance of uniformity of product, the insurance of freshness, and the provision of proper quality control for the consumer. *See* SPICE AND FLAVORING. [A.J.F.]

Soft drinks. These are flavored carbonated waters sweetened with sugar. Cream and root beer sodas are colored with caramel. Cola and ginger ale sodas are colored with caramel and acidified with edible acid. Fruit sodas such as cherry, raspberry, grape, and orange are colored with either natural or certified artificial colors and acidified with fruit acids. In 1957 there were 4819 bottling plants in the United States which sold approximately 32,000,000,000 bottles of soft drinks.

Ingredients of soft drinks require careful treatment and handling. Potable water must be passed through sand and activated carbon filters. Some plants employ chemical flocculation treatment prior to filtration and a few find it necessary to demineralize the bottling water. Processed water is cooled and channeled to enclosed carbonators where it is impregnated with carbon dioxide gas. The pressure of gas insertion and the temperature of the water regulate the amount of gas dissolved. High pressure and low temperature yield high carbonation. Final carbonation in soft drinks is thus controlled.

Carbonated water flows to a bottle-filling machine where it is dispensed into bottles coming from a syruper. Such bottles contain measured volumes of flavored syrup. These bottles have been thoroughly cleaned and sterilized in a mechanized washer.

Flavored syrup consists of predetermined concentrations of sugar, flavoring materials, acid, and color. Flavoring materials, color, and acid are added to sugar solutions (simple syrup). A preservative, sodium benzoate, is used by some bottlers. The flavored syrup is gravitated in sanitary tubing to the syruper of the bottle-filling machine.

Blending of flavored syrup with carbonated water may also be accomplished with syncromix filling systems. Flavored syrup and cooled water are metered into large carbonators and charged with carbon dioxide gas. The blend is then pumped to a filling machine. Filled bottles are sealed by mechanical crimping of special cork-lined metal crowns. They are then conveyed to labelers, casers, and finally to storage or consumer outlet.

In 1957 an estimated 300,000,000 soft drinks in cans were manufactured and sold in the United States. In 1950 dietetic soft drinks appeared on the American market. In these, sugar is replaced with the artificial sweeteners, sodium or calcium cyclamate, which are noncaloric. In 1957 about 400,000,000 bottles were sold. [S.S.E.]

Cereal products. Cereal preparations consist principally of breakfast foods and coffee substitutes. Among the cereal breakfast foods are some

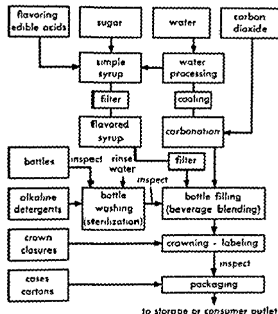


Fig 7. Flow diagram of bottled carbonated beverage manufacture. (American Bottlers of Carbonated Beverages, Washington, D.C.)

that require cooking and some like cornmeal, rolled oats, farina, germ, bran, and cracked wheat, which can be eaten uncooked. The more popular breakfast foods are sold already cooked or require minimal cooking before serving. The manufacture of ready-to-serve breakfast foods is a relatively new industry, developed largely since 1910.

There are two classes of ready-to-serve cereal breakfast foods, those made from entire grains or their milled products, and those made from fabricated cereal products. Both types are found on the consumer market in the form of flaked, puffed, shredded, or granulated breakfast foods. There are considerable variations within each type in regard to the cereals employed, the amount and kind of cooking, and the flavoring used. Flaked products may be made from corn, rice, or wheat. The most popular breakfast food of this type is flaked corn, which is manufactured from corn grits or hominy. Hominy for flaking is cooked in rotary steam cookers for 2–3 hours and is flavored with sugar, salt, malt, honey, or other ingredients while in the cookers. The cooked grits are broken up, cooled, dried to a moisture content of about 15%, and then held in tempering bins for 6–8 hours to equalize moisture distribution. After tempering, the material is passed through heavy-duty flaking machines. The flakes thus produced go to special-type ovens for toasting, after which they are cooled and placed in storage bins until packaged. Vitamins may be added just before packaging.

Puffed-type breakfast foods were among the first developed. Originally only rice and wheat were puffed, but the process has been extended to many other grains such as corn, barley, millet, beans, peas, and soybeans. Puffed breakfast c

also produced from cooked doughs. Puffed cereal products are prepared by placing the material in pressure cookers or puffing guns where the vapor pressure is gradually increased until, when the pressure is suddenly released, the expansion of the water vapor causes an expansion of the product. With wheat and rice, as much as a tenfold volume increase can be obtained. The expansion for other materials is less.

Shredded breakfast food is generally prepared from wheat. Clean wheat is agitated in pressure cookers until the starch is gelatinized completely without disintegrating the kernel. The cooked wheat is cooled and placed in bins for conditioning before going to the shredding machine. Shredding is done by rolls, one of which is smooth while the other has about 20 circular grooves per linear inch cut in the surface. The cooked wheat is forced under pressure between the two rolls which rotate toward each other. This forms the material into slender strands which are carried along on a moving belt where successive layers of strands are added, until the desired thickness has been obtained. The shreds are cut to designated size by flat-edged knives operating from both above and below the flow of material. The cut pieces are placed in baking pans and heated at about 260°C for approximately 20 minutes. After cooling, the product is ready for packaging.

The preparation of granular-type ready-prepared breakfast cereals is a rather complicated procedure. The flowsheet for production of this type of breakfast food is shown in Fig. 8.

A common type of granulated breakfast cereal is prepared from long-extraction wheat flour which is made into a dense, close-grained type of bread by conventional baking methods. The loaves are broken to produce a crumbled product which is again baked and then crushed between coarsely corrugated rolls. The ground material is graded to secure a fairly uniform granulation size and is then packaged for use.

Roasted cereals have been commonly used for a long time as substitutes for coffee. The manufacture of coffee substitutes starts with the careful roasting of whole wheat, which is then granulated and combined with a mixture of roasted bran and sugarcane molasses. The color of the product is used as a guide in judging the length of the roasting period. When the material is removed from the rotary roasters, it is rapidly cooled in pan coolers Products made in this manner are about 50% water-soluble. For the preparation of more water-soluble products, the roasted granulated material is percolated with hot water, the solution clarified, concentrated in vacuo, and finally dried in vacuum drum driers. The dried extract is in brittle sheets which are ground to the desired particle size and packaged for distribution. Besides wheat, roasted barley, rye, and bran are also used for coffee substitutes. [J.A.SH.]

Chewing gum. This is a kind of confection that contains gum base, a masticatory material that does not dissolve in saliva when the gum is chewed. American gum averages in contents 20% gum base, 16% corn syrup, 63% sugar, and less than 1% flavoring such as mint, fruit, and spice.

Chicle, from Central America, was the principal gum base component in 1910. By 1940 jelutong, a more elastic material from Indonesia, was used to a

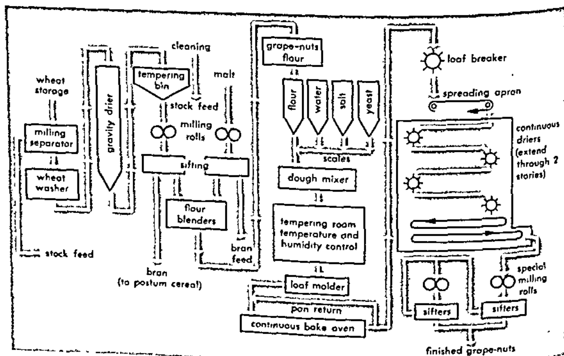


Fig 8. Flowsheet for the manufacture of a granular type cereal breakfast food. (General Foods Corp.)

greater extent. Chicle again became the major component during Japanese occupation of the East Indies, with greatest use in 1944. Since then sorva, from South America, has displaced much chicle and jelutong. All three of these gum bases represent dried latexes from trees. Progressively more and more domestic ingredients such as resins and waxes are replacing the imported products. Calcium carbonate which is a component of most base formulas neutralizes traces of acid forming on the teeth from fermented food residues and so tends to reduce dental caries. Whatever ingredients are used in a gum base, they are blended together hot, and then centrifuged.

The melted gum base is cleaned and mixed with syrup, sugar, and flavor; the resulting chewing gum is shaped by extrusion and rolling. Additional powdered sugar is used on the surface to prevent adhesion to wrappers. A hard sugar surface can be built up on shaped pieces of gum in a tumbling pan, by alternate wetting with a sugar solution and evaporating the added water in a dry air blast. [H.W.C.]

Bibliography: *See* FOOD ENGINEERING.

Food microbiology

The branch of biology that deals with the microorganisms involved in the spoilage, contamination, and preservation of food (*see* FOOD PRESERVATION). All kinds of microorganisms can find their way into food through the ordinary routes of contamination and nearly all of them will find food items suitable media in or on which to grow. Foods handled by people are subject to contamination from all kinds of disease-producing organisms. Food microbiologists must know how to detect the disease-producing organisms, how to prevent their entrance into food, and how to destroy them if they are present. This public health aspect of the subject is very important and involves the use of many special tests and experimental animals.

The extent to which food is contaminated or mishandled is reflected in the number of microorganisms present. Therefore, the food microbiologist must be concerned with the standardization of tests designed to determine the number of organisms and, where possible, the establishment of standards of permissible contamination. With this goes the problem of proper standards of sanitation in the handling of food to avoid unnecessary contamination and growth of microorganisms. It involves an understanding of sanitary engineering practices and equipment design that facilitate adequate cleaning, and a knowledge of how to detect, measure, and reduce contamination.

Another aspect of food microbiology concerns food preservation. Some foods are preserved and desirably flavored by fermentation processes; therefore, the specialist needs to be familiar with the organisms involved as well as their physiology and behavior in fermentation.

Analytical procedures for vitamins and amino acids often use bacteria as assay tools so this activity also becomes a concern of the food microbiologist. [H.O.HA.]

The following sections discuss various aspects of food microbiology concerned with fermented food, fruits and vegetables, eggs, shellfish, meat, and canned foods. *See* BIOASSAY.

Fermented food. There are a number of foods which acquire their special character (flavor, texture, aroma) by a fermentation process involving the participation of bacteria, yeasts, and molds, singly or in combination. Most commonly the microorganisms cause a transformation of carbohydrates (often to lactic acid) and of proteins. Examples of fermented foods are sauerkraut, fermented pickles, Spanish-type olives, certain dairy products (buttermilk, acidophilus milk, yoghurt, kefir, kumiss, leben, and various types of cheese) and certain fermented oriental foods. In the broadest sense fermented alcoholic beverages, vinegar, and yeast-leavened bread may be included, as well as such products as cocoa, coffee, and tea in which microbes participate in the production of the products of commerce. *See* COFFEE; DISTILLED SPIRITS; MALT BEVERAGE; MILK; TEA.

Sauerkraut. Sauerkraut is made by adding about 2.5% salt (weight basis) to shredded cabbage. The salt draws out the plant juice which undergoes a natural lactic acid fermentation, due to the presence of sugars. Usually the heterofermentative *Leuconostoc mesenteroides* develops first, producing lactic and acetic acids, ethanol, mannitol, and carbon dioxide (CO_2). It has a low acid tolerance and when the acid content reaches 0.7–1.0%, it is replaced by the homofermentative *Lactobacillus plantarum*. This organism produces only lactic acid. In the last stages of the fermentation *Lactobacillus brevis* (heterofermentative) completes the process, bringing the total acid content to 1.7–1.8%, of which about one-fifth is acetic acid.

Cucumber pickles. There are many kinds of pickles, some of which are made by a fermentation process. The fermentation of cucumbers is somewhat complex due to the different brining procedures in use. In the manufacture of fermented saltstock an initial brine concentration of about 8% is gradually increased to 16–18% over a period of 6 weeks or more. During this period salt-tolerant lactic acid bacteria (*L. plantarum*, *L. brevis*, and *Pediococcus cerevisiae*) carry out the fermentation. The acid formed is lactic acid and the total amount is usually 0.6–1.0%. When the pickles are packed, much of the salt and acid is leached out and replaced by vinegar, lactic acid, or both, and herbs and spices. In some types of pickles a limited natural fermentation is allowed to proceed in the presence of a low salt concentration (3–5%). Partly fermented dill pickles must be pasteurized to stop the fermentation. They are often referred to as Polish, overnight, or fresh-fermented dill

particular brine treatment depends on the

variety. After the olives are treated with lye to remove the bitter glucoside, they are brined and undergo a lactic acid fermentation. The final salt content is usually 6.5% and the lactic acid content 0.8–1.2%.

Fruits and vegetables. Fruits and vegetables are susceptible to attack by fungi, yeasts, and bacteria. Infection may occur in the field or orchard, in which case plant pathogenic (parasitic) organisms are involved (*see* PLANT DISEASE). After harvest and during storage of the fresh product, parasitic organisms may continue their activity, but in addition saprophytic organisms (those that live on dead organic matter) become active by gaining entrance to the plant tissues through damaged areas. After fruits and vegetables are processed by canning, freezing, or dehydration, or made into modified products (juices, purees, syrups), specific types of saprophytic organisms are encountered as spoilage organisms. *See* BACTERIA; FUNGI; YEAST.

Yeasts and molds are primarily responsible for the spoilage of fruits and fruit products, whereas bacteria are the most common cause of spoilage in processed vegetables. Environmental conditions of storage of fresh as well as processed products determine to a large extent what organisms will develop. These conditions include temperature of storage, presence or absence of oxygen or other gases, humidity, and such intrinsic factors as the pH of the product, its chemical composition and moisture content. In any situation all these factors must be considered interrelated to each other.

Cold storage. During cold storage (above −1 to −0.5°C, the freezing point of most fruits and vegetables) the growth of microbes is merely retarded, but not prevented. Those organisms that grow optimally at temperatures below 20°C (psychrophiles) are especially troublesome. In some products, such as bananas and avocados, physiological chilling injury occurs when they are stored below 45–55°F, and with injury microbial development is accelerated. In other products, such as tomatoes and citrus fruits, storage near the freezing point greatly increases their susceptibility to microbial attack upon return to room temperature. Carbon dioxide gas is sometimes used (5–10%) in the atmosphere of cold storage rooms to retard mold development (especially for apples). Periodic exposure to sulfur dioxide has been found beneficial in the storage of fresh California table grapes. While freezing arrests the growth of spoilage organisms, only a limited percentage of the population is killed, so that upon thawing rapid spoilage of the frost-damaged tissues results.

Drying. Preservation by drying is based on creating a product with high soluble solids content. When drying conditions are poor, sun-drying of fruits may result in the development of osmophilic yeasts and molds. Spoilage of blanched vegetables has been observed in dehydrators when humidity control was neglected and drying was slow. The spoilage is due to starch fermentation by *Aerobacter aerogenes.*

Canning. In canned vegetables thermophilic spore-forming bacteria from the soil are most troublesome. *Bacillus stearothermophilus* and *B. coagulans* are responsible for flat sour spoilage of canned vegetables. *Clostridium nigrificans* causes sulfide stinkers by producing H_2S. Hard swells are caused by the thermophilic *Clostridium thermosaccharolyticum.* It produces CO_2 and hydrogen (H_2). Mesophilic spore-formers occur only in underprocessed canned vegetables or in leaky cans. *Clostridium botulinum* originates in the soil and is a serious cause of food poisoning; outbreaks occasionally occur in home-canned vegetables which have a pH higher than 4.5. [H.J.P.]

Eggs. The interior of the freshly laid egg is usually sterile; subsequent microbial quality is determined by sanitation and storage conditions. No definite pattern characterizes the type of contamination that develops in the interior of eggs, since the microbial flora is chiefly a reflection of the environment. Cracks in the shell and high humidity encourage microbial invasion. Elevated temperatures accelerate spoilage. Organoleptically unsound eggs are called rots; black rots result from anaerobic proteolytic bacteria producing hydrogen sulfide. Molds invade shell eggs under storage at high humidity. Commercial dried eggs if properly processed contain fewer microorganisms than the liquid eggs used as raw materials. *See* EGG PROCESSING.

Shellfish. The microorganisms inhabiting both the surface and interior of shellfish are a reflection of the microbial flora of the water from which they are taken. Contamination by sewage-polluted waters is of major concern. Numerous outbreaks of typhoid fever have been traced to contaminated shellfish. However, holding contaminated shellfish for a time in unpolluted water results in natural purification. Numerous organisms may be responsible for the spoilage of shellfish. The specific type of spoilage is largely determined by the water from which the fish are harvested. The speed of spoilage is determined both by the initial level of organisms and by the storage temperature but occurs fairly rapidly. *See* TYPHOID FEVER.

Meat. Meats are not only stabilized by all common preservation methods, but, in addition, the preparation and merchandizing of many items involves a combination of two or more preservation procedures. Fresh red meats are refrigerated. Cured meats such as ham, bacon, and sausage contain chemical preservatives (salt, nitrate, nitrite) but are, in addition, heat processed and stored under refrigeration. Fresh and cured meats are also canned. With severe heat processing a shelf stable product is produced; certain cured products, such as hams, are canned with only mild heat processing and the resulting product is not shelf stable but must be refrigerated. In certain products stability is attained in part through other processes: fermentation (thuringer), drying (hard sausage), smoking (bacon), and impregnation with vinegar (pig's feet). Each type of meat product has a char-

acteristic microbiological flora; the nature of the flora is determined more by processing methods than by the type of microorganisms which predominate in the raw meat.

Fresh meats. Fresh meats as cut from the chilled animal carcass are contaminated with microorganisms characteristic of the living animal's environment. The living tissues of the animal, with the exception of the gastrointestinal system, are sterile but become contaminated during the slaughtering and dressing operations. While the majority of the contaminants on fresh meats will not grow during refrigerated storage, those that do survive on refrigerated fresh meats consist almost entirely of cold loving bacteria which are able to grow during refrigerated storage. These organisms grow on the meat surfaces and produce alkaline end products.

Cured meats. These are contaminated by acid-producing bacteria, similar to those which sour milk, that do not require oxygen for growth. The bacteria characteristic of fresh meats are apparently excluded from cured products by the presence of salt, nitrate, and nitrite.

Canned meats. Canned meats are either "commercially sterile" (shelf-stable) or carry a microflora consisting of spore-forming bacteria. In perishable canned meats the surviving spores are prevented from growing by refrigeration. In shelf stable cured canned meats the surviving spores are kept dormant through heat shock processing combined with residual curing ingredients.

Canned food spoilage. This may be traced to the growth of microorganisms which survive processing or enter the can through leaks after processing. Spoilage of underprocessed canned foods is almost always due to the growth of heat-resistant spore-forming bacteria. Leaky cans, conversely, show a spoilage microflora reflecting the environment outside the can; usually heat-resistant bacteria [illegible] peas, co[illegible] ilage wherein surviving spores germinate and produce acid but no gas during growth, the unopened can appearing normal. Heat-resistant spores capable of growth at elevated temperatures (above 50°C) may cause swelling of cans of low acid food and the appearance of cheesy or sour odors. In other cases putrefactive anaerobic spore-forming bacteria lead to swelling and putrid odors. In still other cases heat-resistant thermophilic spore-forming bacteria produce hydrogen sulfide (H_2S), which imparts a rotten egg odor to the product.

Acid foods. Such foods as tomato juice and canned fruits receive lower heat treatments and show somewhat different spoilage. Acid-tolerant spore-formers may produce abnormally high amounts of acid without swelling the can. Non-spore-formers, such as yeasts, molds, and lactic acid bacteria, may spoil can contents with or without can swelling.

With the exception of putrefactive spoilage, none of the above abnormalities is a public health hazard. Canned foods are processed in such a manner that chance of contracting botulism, caused by the toxin of a putrefactive bacterium, is lessened. Commercially canned foods have been free of the botulism hazard for over 30 years. *See* BOTULISM. [J.H.SI.]

Bibliography: F. W. Tanner, *The Microbiology of Foods*, 2d ed., 1944; USDA Yearbook of Agriculture, 1950–1951; L. A. Underkofler and R. J. Hickey (eds.), *Industrial Fermentations*, vol. 2, 1954.

Food poisoning, bacterial

Two types are known: the result of (1) infection of the host by organisms carried by food; (2) intoxication by poisonous substances produced by bacteria in food before it is consumed.

Salmonella food poisoning is the most common food-borne infection. Large numbers of these bacteria are necessary to establish human infection; the infection is usually acute, but mortality is very low. Food becomes contaminated by contact with animal or human carriers. *See* SALMONELLA.

Food-borne intoxication is most frequently caused by ingestion of staphylococcus enterotoxin. [illegible] products. The gastrointestinal symptoms are severe, recovery is complete in 3–4 days. *See* STAPHYLOCOCCUS.

Botulism is a rare intoxication usually associated with foods of low acid content (pH above 4.5) canned under insufficiently rigorous conditions. The heat-labile toxin formed during growth of *Clostridium botulinum* acts upon the central nervous system. Mortality averages about 60%. *See* BOTULISM.

Certain other types of food poisoning have been attributed to the growth of streptococci, *Clostridium perfringens*, and *Bacillus cereus*. Insufficient work has been done to determine whether infection or intoxication accounts for the observed symptoms.

Ptomaine poisoning is an archaic and erroneous term once used as a synonym for food poisoning. Ptomaines are diamines found in putrefying material. They are in part responsible for foul odors but their toxicity is low. They have long ceased to be regarded as important in food poisoning. [J.H.SI.]

Food preservation

Processes designed to protect food from spoilage caused by microbes, enzymes, or autooxidation. It may be accomplished by sterilization, refrigeration, desiccation, hermetical packaging, fermentation, chemicals, or by two or more of these in combination.

Sterilization. Sterilization is most readily effected with heat but may also be brought about by radiation and, for some foods, by chemicals. Heat, together with hermetical sealing, is relied upon in

canning. A temperature and time of exposure must be used that are adequate to destroy the heat-resistant bacterial spores. Foods that are not acidic, such as uncured meats and vegetables, will require a temperature of 250°F in the very center of the container for 5–10 minutes. To attain this temperature, steam pressure cookers are needed. Acid foods or salt-cured meats can be preserved by processing at lower temperatures (212°F) for a similar period.

Foods that are sterilized in glass or metal containers must be vacuumized before sealing. This prevents the cans from bursting or seams from opening during processing and produces an indentation in the covers on cooling. To prevent overcooking, the containers should be cooled rapidly after processing, preferably with cold water. During the initial period the seams are expanded and may have pin holes through which the cooling water can enter. Therefore it is necessary to chlorinate this water so that live bacteria are not introduced.

Radiations in the form of either electromagnetic waves (gamma rays) or high-speed electrons (beta rays) are lethal to bacteria and their spores if used in high dosages. For complete sterilization with gamma rays, from 1,000,000–7,000,000 reps (roentgen equivalents physical) are required. Some foods develop off flavors with this high dosage. For these foods, this method becomes impractical with present techniques. Spent rods from atomic piles and radioactive cobalt are suitable sources of gamma rays. These can penetrate metal and glass containers so that the food can be sterilized in the container, as is done with steam sterilization.

Refrigeration. Refrigeration is a reliable method of preservation but is more effective at temperatures below freezing than above. In common usage the term refrigerated foods designates those preserved at temperatures just above freezing, and frozen foods, those preserved at lower temperatures.

Most spoilage bacteria grow very slowly at temperatures below 7.2°C and some do not grow at all. A satisfactory refrigerator should therefore be kept below this temperature—the closer to 0°C the better. At best, foods cannot be preserved in the refrigerator more than 1 or 2 weeks.

Frozen foods will keep for long periods. Although bacteria are not killed by the freezing, they remain dormant in the frozen state unless the temperature is quite close to the freezing point. Enzymes that have been elaborated by bacteria or that are present in the food can function slowly; however, to stop these from acting, it is best to destroy them by blanching, that is, by heating momentarily to near the boiling point before [illegible]
Some fo[illegible]
growth c[illegible]
temperature should be kept constant.

Dehydration. This is nature's method of preserving seeds and other vital living systems from one season to another. Microorganisms of all kinds require moisture for growth, therefore growth can be stopped by reducing the moisture level below certain critical levels. This level varies with the organism and with the ability of the item to bind the water so that it is not present as free moisture. The actual moisture content is therefore not as important as the equilibrium vapor pressure (evp) of the product. For example, a product with an evp of 75 would neither gain or lose water in an atmosphere where the vapor pressure was 75% of the vapor pressure of pure water held at the temperature of the product.

Most bacteria cannot grow if the evp is below 94, while molds can grow to an evp of 75. Summer sausage will reach this critical level for bacteria when the moisture content is about 34%; for grains it must be as low as 15% to reach the same critical level. Substances such as salt and sugar may be added to the food item to help lower the evp and thus preserve it at a higher moisture content. Thus, summer sausage owes its keeping quality to partial drying and to the binding of water by salt and sugar.

Use of chemicals. Chemical preservation is possible with a variety of substances such as common salt, nitrate, and nitrite, various acids and alcohol, chemicals from smoke, and antibiotics. In all such applications it must be shown that the chemical used is not injurious to health nor deleterious to the nutritive value. Chemicals that have been used for this purpose through the ages are permissible, but any new ones proposed must be thoroughly tested and are subject to Federal regulation.

Salt. Salt is effective in stopping the growth of most common spoilage bacteria in concentrations of 5% or higher, calculated on the basis of water present in the food. It is primarily responsible for the keeping quality of salt-cured meats, fish, and sausage.

Nitrates and nitrites. These chemical compounds have been used in the curing of meat since ancient times, and while they are added primarily for the purpose of preserving the color, they also inhibit the growth of some anaerobic bacteria and thus help the salt in the preservation of meat. These chemicals are somewhat toxic. The amount that can be used must therefore be limited to a few ounces of nitrate and a fraction of an ounce of nitrite per 100 pounds of meat.

Organic acids. The most effective and commonly used preservatives among these are benzoic, sorbic, acetic, lactic, and propionic acid. They may be added in the form of their sodium or potassium salts, but to be effective the product must be somewhat acidic. Benzoic acid should not be used in excess of 0.1%. It is found naturally in cranberries and loganberries and is responsible for their excellent keeping quality.

Sorbic acid and propionic acid used in concentrations of not more than 0.3% are very effective for the prevention of mold growth. Acetic, lactic, and propionic acids are products of bacterial me-

tabolism and thus food can be preserved by fermentation. Enough of the acids should be added or formed to produce a pH of 5.0 or less. They are responsible for the preservation of sauerkraut, pickles, and fermented milk.

Smoking. Smoking, like salt curing, has been used from ancient times. Its effect is exerted by com- [illegible] he surface [illegible] : preserva- [illegible]

[illegible] Antibiotics have attracted the attention of food microbiologists and manufacturers. If an antibiotic could be found that would stop the growth of all food spoilage organisms and at the same time be harmless for humans, it would be an ideal [illegible] h an [illegible] The [illegible] ould not be used unless they are completely destroyed in the cooking of the food so as not to build up a population of disease-producing bacteria resistant to them. To date, chlortetracycline and oxytetracycline have been sanctioned by the Federal government for use in fresh fish and poultry. These antibiotics enhance the keeping quality of these products when used in conjunction with procedures such as refrigeration and adequate sanitation that are customarily necessary for the preservation of the products. See ANTIBIOTIC; CHLORTETRACYCLINE; OXYTETRACYCLINE. [H.O.HA.]

Food science

The application of such basic sciences as chemistry, bacteriology, microbiology, and mathematics to the study of foods. These studies include food from field through harvesting and processing to the table, and subsequent phases of ingestion and utilization by man.

Food science [illegible] fo [illegible] ta [illegible] in [illegible] re[illegible] and analytical work.

Food science is taught in over 30 American universities and colleges where a major in food science, food technology, or food engineering is offered. Degrees available include B.S., M.S., and Ph.D. Universities prominent in teaching the subject include Massachusetts Institute of Technology, Ohio State University, Oregon State College, Rutgers University, University of Illinois, and University of California. Several schools also teach specialized phases of food science, such as dairy technology and cereal chemistry.

The official professional organization in the field is Institute of Food Technologists, with about 6000 members in the United States and abroad. Phi Tau Sigma Honorary Society encourages the application of fundamental scientific principles to food science; it honors and recognizes professional achievement in the field.

The IFT publishes two official organs, *Food Technology*, where research articles of commercial interest and news of the society are published; and *Food Research*, in which the more scientific articles on basic research are presented.

Persons with food science training are engaged in all lines of endeavor related to food production, preservation, manufacturing, distribution, and consumption. Many are employed in teaching and research in universities; in federal, state, and municipal inspection and regulatory agencies; in U.S. Government research laboratories, and in research product development, quality control, sanitation, production, packaging, and sales operations of commercial firms producing manufactured and processed foods, food machinery, food ingredients, and packaging materials and containers. See FOOD ENGINEERING; FOOD MICROBIOLOGY; FOOD TECHNOLOGY. [R.E.M.]

Food spoilage

Foods undergo many different types of spoilage, resulting in abnormal colors, flavors, odors, or in other changes, such as consistency. Spoilage cannot be equated to danger to the consumer; indeed, in most instances, foods which contain poison of either chemical or biological origin show no external evidence of spoilage.

Perishable foods. Fresh meat, vegetables, and fruits have a limited shelf life, after which time spoilage is expected, even though all phases of processing and storage may be optimal. The shelf life of perishable foods is determined by the initial quality of the raw materials, the method of processing, and subsequent handling.

Spoilage of perishable products is usually characteristic, being determined primarily by the chemical nature of the food, the type of microbial contamination it carries, its packaging, and its storage condition. For instance, raw milk sours because of the activity of lactic acid forming bacteria. Fresh meats develop characteristic slime and off odors as a result of the growth of bacteria which produced alkaline conditions. Frankfurters and certain other cured meat products develop a surface slime due to the growth of acid forming bacteria similar to those found in dairy products.

Shelf stable foods. Canned and dried foods are processed with the aim of producing a food which may be stored for indefinite periods even at warm temperatures without undergoing apparent change. Stability is imposed by such factors as heat (canned foods), drying (dried vegetables), and chemicals (addition of acid, such as in mayonnaise).

Spoilage of shelf stable foods is, by definition, an abnormal and unexpected condition. Underprocessing of canned food leads to unstable products in which the contents of the can undergo putrefaction or fermentation with the development of abnormal odors, consistencies, and flavors. Usually, the instability of the product is evinced by the swelling of the unopened can. Insufficient dehydration of dried food products or improper storage in humid atmospheres similarly leads to off condi-

tions in dried food products. Failure to add sufficient vinegar to mayonnaise renders the product subject to microbial attack.

Spoilage mechanism. Chemical changes are responsible for spoilage. These changes are brought about through enzyme action. These enzymes may have their origin in the food material or may be elaborated by microorganisms such as yeasts, molds, and bacteria which contaminate the product. In general, enzymes derived from foods themselves are of secondary importance to those elaborated by microorganisms; however, when foods are held under conditions which preclude microbial action, the natural enzymes in the food materials may be the primary reason for spoilage, for example, fat splitting enzymes in frozen foods.

Yeast and mold spoilage. This is most characteristic of foods with high acid or sugar content. Molds characteristically grow on the surface; yeasts which are capable of growth in the absence of air may spoil food either by the formation of slime or by fermentation of the entire food mass. Molds occasionally present a spoilage problem in cream, eggs, various sugar products, butter, and cheese. Yeasts are associated with the fermentation of candy, butter, mayonnaise, catsup, syrup, olives, and pickles. In certain foods, the addition of benzoic, propionic, and sorbic acids aids in the control of yeasts and molds.

Bacterial spoilage. This type of food product spoilage is as varied as the capabilities of this large group of microorganisms. The major classes of food substances are all subject to some form of bacterial decomposition, and within each class, there is frequently a major type or pattern of spoilage which is characteristic. The most important determinants with respect to the type of spoilage are the methods of processing and storage. For example, the spoilage of fresh meats is primarily a surface phenomenon involving highly aerobic bacteria. If, however, fresh meats are stored in the absence of oxygen, they will undergo spoilage of a different type, a spoilage brought about by lactic acid forming bacteria of the same general characteristics as those which spoil fresh milk. The same fresh meat product, if canned, may undergo spoilage, but the bacteria involved are heat-resistant forms which survive processing, the so-called spore-forming bacteria. By analogy, if fresh milk were aerated the same organisms which spoil fresh meats could be expected to spoil the fresh milk; similarly, spore-forming organisms are involved in the spoilage of canned milk products.

Abnormal or unusual spoilages. Abnormal or unusual spoilages of food products occur periodically. Such spoilage may be seasonal, regional, or entirely haphazard in appearance. In extreme cases a food plant may have to cease manufacture of a product due to contamination with undesirable microorganisms responsible for peculiar spoilage problems. Examples of such unusual spoilage include green discoloration of milk; gas pockets and, in extreme cases, the explosion, of cheese. In most cases, a particular organism or group is responsible for each of these conditions. *See* FOOD ENGINEERING; FOOD MICROBIOLOGY; FOOD PRESERVATION. [J.H.SI.]

Food technology

[illegible] to the

[illegible]

neering. However, many food technologists by training and experience are food scientists rather than engineers. *See* FOOD ENGINEERING.

As defined by the Institute of Food Technologists, "food technology is primarily based on the fundamentals of chemistry, physics, biology, and microbiology, any of which sciences may find an expression through an engineering operation.

"Knowledge of food technology enables its possessor to develop new products, processes, and equipment, to select proper raw materials, to understand and control manufacturing operations, to solve technical problems of food manufacture and distribution, including those involved in plant sanitation, and those affecting the nutritional value and public health safety of foods, and to know the fundamental changes of composition and of physical condition of foodstuffs which may occur during and subsequent to the industrial processing of the foodstuffs."

Degrees in food technology are given by more than a score of leading colleges and universities. However, the amount of mathematics, physics, and engineering principles included in the curriculums varies considerably. That is why some food technologists are trained as food engineers, but others are primarily food scientists. The former are best qualified for food manufacturing and processing operations, the latter for research and quality control.

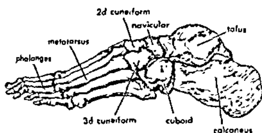
Specialized branches of food technology, science, and engineering are taught in many institutions of higher education, primarily state colleges and universities with agricultural schools. Thus, it is possible to obtain degrees in dairy, cereal, fruit and vegetable, baking, meat, and brewing chemistry or technology.

Specialized phases of food technology, such as food chemistry, bacteriology, and nutrition are taught in many schools with scientific curriculums.

To the extent that they are applied in developing and growing suitable raw materials for manufacture and processing, horticulture and animal husbandry are food sciences. [F.K.L.]

Foot

The terminal portion of the lower extremity. In man, the bony framework consists of 7 tarsal, or ankle bones, 5 metatarsals, or foot bones, and 14 phalanges, or toe bones. The prominent heel is the projection of the calcaneus, a tarsal *bone*. A strong but elastic system of ligaments holds the bones in position and strengthens the joints of the foot.



Foot, lateral aspect. (From W. H. Lewis, ed., *Gray's Anatomy of the Human Body*, 22d ed., Lea and Febiger, 1930)

Ankle, foot, and toe movements are produced by the action of 3 groups of leg muscles and 6 groups of foot muscles. The major blood supply and drainage is through the tibial arteries and saphenous veins, respectively. An extensive lymphatic system is present, particularly in the rather thick subcutaneous tissues. Motor and sensory nerves are supplied by branches from the lumbosacral plexus Homologous and analogous structures are present in all tetrapodal vertebrates, but wide variations exist, depending upon class features and individual species adaptation. *See* MUSCULAR SYSTEM; SKELETAL SYSTEM. [E.C.ST.]

Foot (unit)

A unit of length employed in English-speaking countries. On the basis of an agreement made in 1959 by the directors of the standards laboratories of the English-speaking countries, 1 foot is defined as exactly 0.3048 meters. The foot is the length unit used in the British systems of units employed in engineering practice. *See* UNITS, SYSTEMS OF; *see also* METER (UNIT). A yard is defined as exactly 3 ft, or 0.9144 m. [D.WI.]

Foot and mouth disease

A highly contagious virus disease of economic importance, infecting cattle, pigs, sheep, and goats. It is characterized by fever, salivation, and formation of vesicles in the mouth and pharynx and on the feet. It may be transmitted to man.

Epizootics may be characterized by high mortality rates. Survivors are poor producers of milk and meat. The infective virus is about 22 millimicrons in diameter. A variety of experimental animals and tissue cultures can be infected.

Diagnosis is by isolation of virus and by serologic tests. *See* SEROLOGY.

Formalin-inactivated virus vaccines immunize, but only temporarily. In serious epizootics, all susceptible animals in the area are slaughtered, and strict quarantine is observed. To prevent human infection, boiling or pasteurizing of farm products under suspicion should be adequate. *See* ANIMAL VIRUS; EPIDEMIOLOGY. [J.L.M.]

Foot-candle

The unit of illumination when the square foot is taken as the unit of area. It is the illumination on a surface, 1 ft² in area, on which there is a luminous flux of 1 lumen uniformly distributed, or more simply, a density of 1 lumen/ft² falling on a surface. It may also be defined as the illumination on a surface all points of which are at a distance of 1 ft from a uniform point source of 1 candle.

The foot-candle is the unit of illumination in general practical use in the United States at present. *See* CANDLE; ILLUMINATION. [R.C.P.]

Foot-lambert

A unit of luminance (photometric brightness) equal to $1/\pi$ candle per square foot. Since there are 144 square inches in a square foot, the luminance in candles/in.² multiplied by 144π (452) gives the luminance in foot-lamberts.

The foot-lambert is also defined as the uniform luminance of a perfectly diffusing surface emitting or reflecting light at the rate of 1 lumen/ft². Since the foot-candle is defined as 1 lumen/ft² falling on a surface, the foot-lambert is sometimes referred to as the apparent foot-candle. The average luminance of a surface in foot-lamberts is, therefore, the product of the reflection factor of the surface multiplied by the illumination in foot-candles. *See* FOOT-CANDLE; LUMEN; LUMINANCE. [R.C.P.]

Foot-pound

A unit in the English gravitational system, having two definitions. (1) A unit of energy or work, being equal to the work done by 1 pound of force (1 lbf) when the point at which the force is applied is displaced 1 ft in the direction of the force; the abbreviation is ft-lbf. *See* WORK.

(2) A unit of torque, being equal to the time-rate of change of angular momentum produced by a pound of force (lbf) acting at a perpendicular distance of 1 ft from the axis of rotation; it is sometimes called pound-foot (lbf-ft), to distinguish it from the unit of energy or work. *See* TORQUE; *see also* UNITS, SYSTEMS OF. [D.E.R.]

Foot-poundal

A unit in the English absolute system, having two definitions. (1) A unit of energy or work, being equal to the work done by a force of magnitude 1 poundal when the point at which the force is applied is displaced 1 ft in the direction of the force; the abbreviation is ft-pdl. *See* WORK.

(2) A unit of torque, being equal to the time-rate of change of angular momentum produced by a force of magnitude 1 poundal acting at a perpendicular distance of 1 foot from the axis of rotation; it is sometimes called poundal-foot (pdl-ft), to distinguish it from the unit of energy or work. *See* TORQUE; *see also* UNITS, SYSTEMS OF. [D.E.R.]

Foraminifera fossils

Foraminifera constitute an order of protozoans of the class Sarcodina. Most foraminifers are small, the test (shell) being less than 1 mm in diameter. A few species may have tests as large 100 mm. The order contains a number of forms which are predominantly marine. As

foraminifers are widely distributed in marine sediments and in rocks which were laid down in ancient seas. They vary in abundance in different strata and are useful in stratigraphic correlation, particularly in determining oil-bearing deposits. Many forms obtained in deep-sea cores are used as indicators of climatic change during glacial and postglacial time. *See* MARINE SEDIMENTS; SARCODINA.

Foraminiferal tests. The preservable hard parts of foraminiferal tests are composed primarily of chitinous, arenaceous, siliceous, and calcareous materials.

Chitinous. Tests of this substance are the simplest and possibly the oldest type in form, consisting of a thin dark or transparent flexible layer of chitinous (tectine or keratine) material. Chitinous linings are observed in some fossil genera of the arenaceous (sandy), perforate and imperforate, calcareous group of Foraminifera. Chitinous tests prevail in Recent Allogromiidae and are common in living brackish-water Foraminifera. They are rarely preserved as fossils.

Arenaceous. These tests, commonly referred to as agglutinated or adventitious tests, are formed almost entirely by agglutination of foreign mate- [illegible] basal layer and firmly held together with ferruginous, calcareous, or siliceous cement. The tests are commonly insoluble in weak acid. The oldest known fossil Foraminifera, of superfamilies Astrorhizidea and Lituolidea, have this type of shell material.

Calcareous. This material occurs in two main types, imperforate calcareous, consisting of cryptocrystalline calcite without pores and of porcellaneous appearance (superfamily Miliolidea); and perforate calcareous with a translucent (hyaline) appearance and variable size of perforations or [illegible] superfamilies Nodo- [illegible] have aragonitic tests.

Ecology. Each basin of sedimentary deposition throughout geological time is characterized by a variety of environments or habitats. These range from shallow nearshore to deeper offshore environments in both warmer and colder regions. The ecological relationships of such environments are often reflected in the over-all composition of the fossil foraminiferal assemblages which are embedded in the various types of sediments. Paleoecological analyses of assemblages of bottom living (benthonic) and open-sea (pelagic or planktonic) Foraminifera, based on comparisons with Recent associations living under known conditions, provide an indication of the depositional depth and of the temperature and salinity which prevailed at the time of deposition of the embedding sediment.

A large number of the facies maps and paleogeographic maps which are drawn are based mainly on benthonic Foraminifera. More than 98% of all Foraminifera are benthonic and some of them are very sensitive to changes in environmental conditions. Planktonic Foraminifera which can be swept

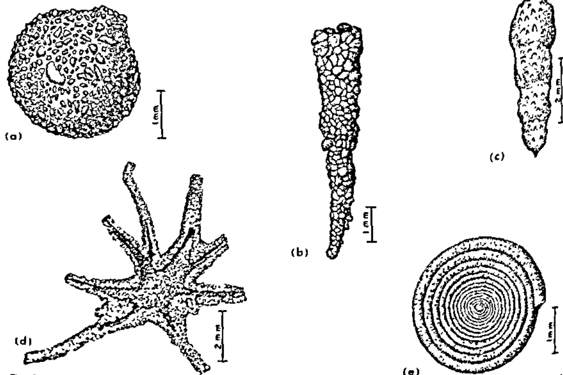


Fig. 1. Astrorhizidea. (a) *Saccammina sphaerica* Sars. (b) *Jaculella acuta* Brady. (c) *Haplostiche constricta* (Reuss). (d) *Astrorhiza limicola* Sandahl, showing interior. (e) *Ammodiscus incertus* (Orbigny).

Characteristics of Foraminifera superfamilies

Superfamily	Material of shell wall	Structure and shape of shell	Aperture	Range
Astrorhizidea	Arenaceous, agglutinated	Tubular, branching stellate, subspherical nonseptate, irregularly constricted	Simple	Cambrian to Recent
Lituolidea	Agglutinated, arenaceous; advanced forms labyrinthic; granular calcareous	Straight, uniserial, biserial, planispiral, trochospiral, multilocular	Simple or complex	Devonian to Recent
Fusulinidea	Arenaceous, composite, granular or fibrous calcareous in 1–4 different layers	Coiled, typically planispiral, lenticular, fusiform, globular complex septation multilocular	Basal	Devonian to Recent
Miliolidea	Calcareous imperforate (porcellaneous) adventitious material on external layer	Coiled in varying planes, or planispiral, multilocular	Simple, terminal; single, dentate, or cribrate	Carboniferous to Recent
Nodosariidea	Calcareous perforate (hyaline), usually ornamented	Uniserial, straight, curved, planispirally coiled; uncoiling or arranged about a longitudinal axis, multilocular	Terminal, radiate, or simple slit or round opening	Devonian to Recent
Buliminidea	Calcareous perforate, usually ornamented	Trochospiral to biserial and straight, uniserial, multichambered	Basal or terminal, slitlike often with tooth or plate or tube	Upper Triassic to Recent
Spirillinidea	Calcareous perforate; consisting of calcite crystal	Planispiral or trochospiral, long tubelike chamber unilocular	Open tube ending or at base of ventral side	Jurassic to Recent
Rotaliidea	Calcareous perforate, radial arrangement, usually ornamented	Trochospiral, planispiral, annular or cyclic, complex; multichambered, multiseptate	Slit along basal suture, multiple or complex	Triassic to Recent

into any environment by currents are known to have existed since the beginning of the Cretaceous Period and are of great value, owing to their worldwide distribution, for interbasinal and intercontinental stratigraphic correlations. Of the 60 known families only 4 (Hantkeninidae, Orbulinidae, Globorotaliidae, and Globotruncanidae) are of the planktonic mode of life. Most genera in these families show definite evolutionary trends and are known only as fossils. *See* FACIES (GEOLOGY); PALEOECOLOGY; PALEOGEOGRAPHY.

Throughout geological time certain genera were substantial rock builders. Examples are *Fusulina* during Carboniferous and Permian; orbitoidal reef-builders during pre-Pliocene Tertiary; *Nummulites* and *Alveolina* during the Eocene; *Globotruncana*, during Upper Cretaceous; *Globigerina*, *Amphistegina*, and *Operculina* during Miocene-Pliocene; and miliolids as oil reservoir rocks during the Middle Cretaceous in Mexico and during the Oligocene in Iran. Tertiary marls composed of *Globigerina* have their counterpart in Recent *Globigerina* ooze which covers more than one-half of the bottom area in the Atlantic and Indian Oceans. *See* LIMESTONE.

[illegible] superfamilies and gen- [illegible]

primitive agglutinating and geologically oldest genera which are distributed in 9 families. Some of the genera are excellent ecologic indicators and interesting from an evolutionary point of view (Fig. 1). [illegible] for age determina- [illegible]

which some fossil genera are stratigraphically very important, such as *Choffatella* (Jurassic-Lower Cretaceous); *Loftusia* (uppermost Cretaceous), with specimens as large as 12 cm long; some textulariids and valvulinids; the whole family of the conically shaped Orbitolinidae (Lower Cretaceous); and the spirally coiled Endothyridae (Carboniferous), Fig. 2. Others are excellent ecologic indicators, for instance, the trochospiral Trochamminidae which can live in waters of very low salinity (brackish water).

Fusulinidea. Members of this group, which is represented by two families (Fusulinidea, Neoschwagerinidea) are stratigraphically of

Fig. 2 Lituolidea (*a*) *Haplophragmoides elegans* (Grzybowski). (*b*) *Ammomarginulina bellensis* Loeblich. (c) *Cyclammina cancellata* Brady. (*d*) *Textularia gramen* Orbigny. (e) *Martinottiella communis* (Orbigny). (*f*) *Trochammina inflata* (Montagu). (*g*) *Triplasia goodlandensis* Cushman and Alexander. (*h*) *Endothyranella armstrongi* Plummer.

Fig. 3. Miliolidea. (*a*) *Miliolina* (*Quinqueloculina*) *auberiana* (Orbigny). (*b*) *Articulina sagra* Orbigny. (*c*) *Spiroloculina canaliculata* Orbigny. (*d*) *Peneroplis pertusus* (Forskal). (*e*) *Sorites dominicensis* Ehrenberg (*f*) *Alveolina*, showing interior structure.

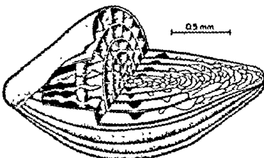


Fig. 4. Fusulinidea; *Triticites*, showing internal structure.

groups of fossil Foraminifera from a stratigraphic, phylogenetic, and evolutionary point of view. *See* Fusulinidae.

Miliolidea. This group is composed of five families. The most important are the Alveolinidae, which include genera with involute planispirally coiled, globular to cylindrical forms that are as much as several centimeters in length. They have an intricate structure of the test similar to the Paleozoic fusulinids, although they are differently built (Fig. 3). The Miliolidae, like the Alveolinidae, also are often rock formers. Another family, Peneroplidae, includes genera with a fan-shaped test of annular rows subdivided into minute radial chamberlets.

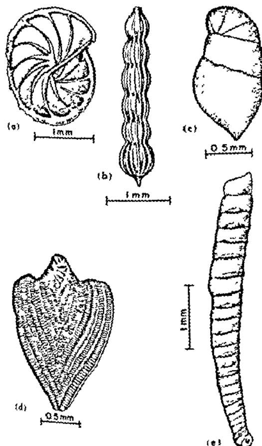


Fig. 5. Nodosariidea. (a) *Robulus navarroensis* (Plummer). (b) *Nodosaria affinis* Reuss. (c) *Marginulina glabra* Orbigny. (d) *Palmula reticulata* Orbigny. (e) *Vaginulina cretacea* Plummer.

importance. More than 80 genera are known, some of which have a highly complex and intricate shell structure. Many of the genera had short stratigraphic ranges and were prominent rock builders. They are found in Upper Carboniferous and Permian rocks of all continents. None survived the end of the Permian. Although most genera were only a few millimeters or at most about 1–1.5 cm long, some specimens attained a length of 10.2 cm (Fig. 4). The fusulinids are one of the best studied

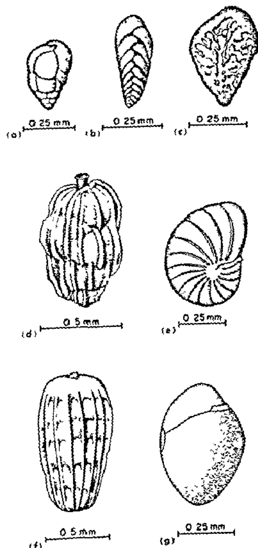


Fig. 6. Buliminidea. (a) *Bulimina exigua* Cushman and Parker. (b) *Bolivina cretosa* Cushman. (c) *Bolivinoides decoratus* (Jones). (d) *Uvigerina gallowayi* Cushman. (e) *Nonion costiferus* (Cushman). (f) [illegible] ...edi Cushman. (g) *Chilostomella ovoidea* Reuss.

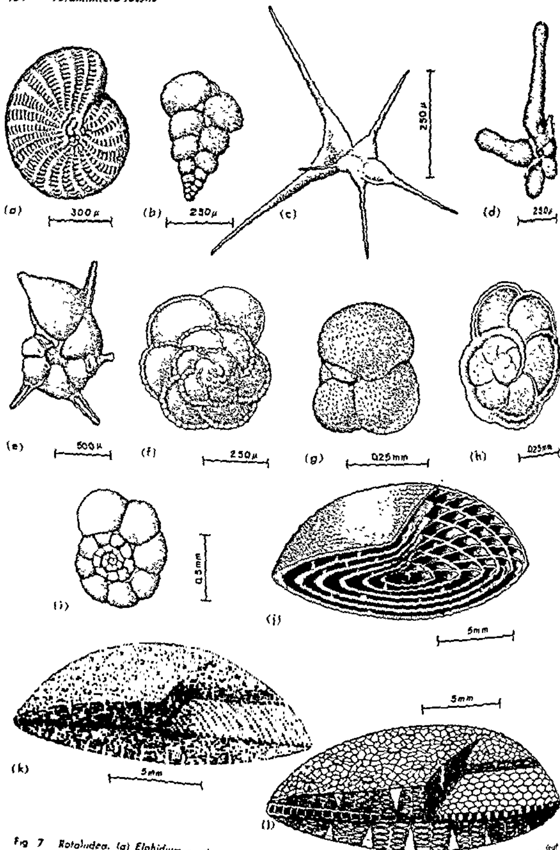


Fig 7 Rotaludea. (a) *Elphidium arcticum* (Orbigny). (b) *Gümbelina planata* Cushman. (c) *Schackoina maltiensis* Reichel (d) *Hastigerinella watersi* Cushman. (e) *Hantkenina primitiva* Cushman and Jarvis. (f) *Globotruncana arca* (Cushman). (g) *Globigerina finaperta* Finlay. (h) *Globorotalia menardii* (Orbigny). (i) *Valvulineria californica* Cushman. (j) Diagram (cut open) of one-half of the test of *Nummulites* (k) Diagram (cut open) of one-half of the test of *Discocyclina* (l) Diagram (cut open) of one-half of the test of *Lepidocyclina*.

Nodosaridea (Lagenidea). Members of this group, although known from Devonian to Recent, became prominent especially during Jurassic and Early Cretaceous time. They include three families (of which the Nodosariidae are the most important), contain the greatest number of genera, and exhibit a great variety of shapes in their often intricately ornamented perforate calcareous tests (Fig. 5). Polymorphism characterizes the test in the family Polymorphinidae, which somewhat resembles the imperforate calcareous Miliolidae in the spirally arranged chambers of the tests.

Buliminidea. Five families are included in this group, in which many genera also exhibit a great variety of shape of the test. This is shown by *Reussella* with a [illegible]

[illegible]

rotating chamber arrangement in *Cassidulina*, and the planispiral involution in *Nonion* (Fig. 6).

Spirillinidea. This superfamily is represented by only one family, with a rather insignificant and stratigraphically unimportant group of genera. These include forms which are either planispirally (*Spirillina*) or trochoidally coiled (*Patellina*) and followed by a long undivided chamber tube, the whole test being formed by one or a few large calcite crystals.

Rotaliidea. This large superfamily includes structurally complex, highly developed, polymorphic genera brought together in 25 different families, some of which are known only from fossils, for example, Schackoinidae (Upper Cretaceous), Hantkeninidae (Eocene), Heterohelicidae (Cretaceous to Oligocene), Miscellaneidae (Upper Cretaceous to Upper Eocene), the subfamily Nummulitinae (Eocene to Oligocene), Orbitoididae, Pseudorbitoididae, and Lepidorbitoididae (Upper Cretaceous), Discocyclinidae (Eocene), Asterocyclinidae (Eocene), Helicolepidinidae (Middle and Upper Eocene), Lepidocyclinidae (Eocene to Lower Miocene), and Miogypsinidae (Middle Oligocene to Lower Miocene). Four other families include exclusively planktonic genera. Many genera of the different families are excellent geological time markers (*Globotruncana, Hantkenina, Globorotalia*, and *Globigerina*). Others are ecologically important, or are substantial rock-builders, as is the case with most of the reef-dwelling and extinct *Nummulites, Lepidocyclina, Discocyclina, Miogypsina*, and the Orbitoididae (Fig. 7). The main development of this superfamily occurred during Upper Cretaceous and Tertiary times. *See* NUMMULITES.

The Recent superfamily Allogromiidea with its chitinous, tectinous, or keratinous tests, although possibly the ancestral stock of the Foraminifera, has not left any traces in the geologic record and therefore has been omitted from the accompanying table.

Uses. Although there were some scattered attempts made during the nineteenth century, and in spite of the fact that the first fossil Foraminifera were studied more than 300 years ago, the practical use of this group of microfossils as indicators for age and ecology of marine sediments dates only from the end of World War I. Applied foraminiferology is clearly a product of the oil industry which on account of an accelerated intensive search for hidden crude-oil and gas resources since 1920 had to utilize microfossils, in absence of determinable megafossils, for the relative age determination of the sediments pierced in oil-well drilling. Stratigraphic correlation may be from well to well, between the wells within an oil field, or between oil fields of the same and adjoining oil basins. Finally, the intercontinental synchronization of formations is possible with the help of the ubiquitous planktonic species of fossil Foraminifera. This is one of the main tasks of every oil-company micropaleontologist, and hand in hand with it goes a paleoecological interpretation of fossil foraminiferal assemblages so important for localizing stratigraphic oil traps and for unraveling many paleogeographic problems such as extension of former seas, distribution of sediments of different bathymetric depths, and facies changes in rock formations. *See* MICROPALEONTOLOGY; PROSPECTING, PETROLEUM; *see also* INDEX FOSSIL; STRATIGRAPHY.

[H.E.T.]

Bibliography: J. A. Cushman, *Foraminifera, Their Classification and Economic Use*, 1948; J. J. Galloway, *A Manual of Foraminifera*, 1933; F. M. Glaessner, *Principles of Micropaleontology*, 1947; J. Piveteau, *Traité de Paléontologie*, vol. 1, 1952; V. Pokorný, *Grundzüge der zoologischen Mikropaläontologie*, vol. 1, 1958.

Foraminiferida

An order of the class Rhizopodea also known as the Foraminifera. These Protozoa have myxopodia (rhizopodia) and tests, often multichambered, typically enclosed in cytoplasm. Living species are mostly marine and brackish water types. A few are pelagic, the rest creeping or sessile. Calcareous and siliceous tests are composed of secreted materials. Arenaceous tests consist mainly of foreign particles cemented together. During growth of multilocular types, chambers are added to an initial test, the proloculum, in various patterns. Tests measure from less than 1.0 mm to about 50 mm (*Bathysiphon*) and, in such fossils as *Camerina* (*Nummulites*), up to about 100 mm. Foraminiferida are mainly bottom dwellers. Their tests accumulate in deposits like Globigerina ooze which eventually become sedimentary rock. These fossils are useful in determining geological correlation.

[R.P.H.]

Force

Force may be briefly described as that influence on a body which causes it to accelerate. In this way force is defined through Newton's second [illegible] motion.

This law states in part that the [illegible] body is proportional to the res[illegible]

erted on the body and is inversely proportional to the mass of the body. An alternative procedure is to try to formulate a definition in terms of a standard force, for example, that necessary to stretch a particular spring a certain amount, or the gravitational attraction which the earth exerts on a standard object. Even so, Newton's second law inextricably links mass and force. *See* ACCELERATION; MASS.

Many elementary books in physics seem to expect the beginning student to bring to his study the same kind of intuitive notion concerning force which Sir Isaac Newton possessed. One readily thinks of an object's weight, or of pushing it or pulling it, and from this he gains a "feeling" for force. Such intuition, while undeniably helpful, is hardly an adequate foundation for the quantitative science of mechanics.

Newton's dilemma in logic, which did not trouble him greatly, was that in stating his second law as a relation between certain physical quantities, he presumably needed to begin with their definitions. But he did not actually have definitions of both mass and force which were independent of the second law. The procedure which today seems most free of pitfalls in logic is in fact to use Newton's second law as a defining relation

First one supposes length to be defined in terms of the distance between marks on a standard object, or perhaps in terms of the wavelength of a particular spectral line. Time can be supposed similarly related to the period of a standard motion (for example, the rotation of the earth about the sun, the oscillations of the balance wheel of a clock, or perhaps a particular vibration of a molecule) Although applying these definitions to actual measurements may be a practical matter requiring some effort, a reasonably logical definition of velocity and acceleration, as the first and second *time derivatives* of vector displacement, follows readily in principle.

Absolute standards. Having chosen a unit for length and a unit for time, one may then select a standard particle or object. At this juncture one may choose either the absolute or the gravitational approach. In the so-called absolute systems of units, it is said that the standard object has a mass of one unit. Then the second law of Newton defines unit force as that force which gives unit acceleration to the unit mass. Any other mass may in principle be compared with the standard mass by subjecting it to unit force and measuring the acceleration, with which it varies inversely. By *suitable appeal to experiment*, it is possible to conclude that masses are scalar quantities and that *forces are vector quantities* which may be superimposed or resolved by the rules of vector addition and resolution.

In the absolute scheme, then, it is written for nonrelativistic mechanics that

$$\mathbf{F} = m\mathbf{a}$$

where boldface type denotes vector quantities. The quantities on the right are previously known, and this statement of the second law of Newton is in fact the definition of force. In the absolute system, mass is taken as a fundamental quantity and force is a derived unit of dimensions MLT^{-2} (M = mass, L = length, T = time).

Gravitational standards. The gravitational systems of units regard the attraction of the earth for the standard object as the standard force. Newton's second law still couples force and mass, but since force is here taken as the fundamental quantity, mass becomes the derived factor of proportionality between force and the acceleration it produces. In particular, the standard force (the earth's attraction for the standard object) produces in free fall what one measures as the gravitational acceleration, a vector quantity proportional to the standard force (weight) for any object. It follows from the use of Newton's second law as a defining relation that the mass of that object is $m = w/g$, g being the magnitude of the gravitational acceleration and w being the magnitude of the weight. The derived quantity mass has dimensions FT^2L^{-1}.

Because the gravitational acceleration varies slightly over the surface of the earth, it may be objected that the force standard will also vary. This may be avoided by specifying a point on the earth's surface at which the standard object has standard weight. In principle, then, the gravitational system becomes no less absolute than the so-called absolute system.

Composition of forces. By experiment one finds that two forces of, for example, 3 units and 4 units acting at right angles to one another at point 0 produce an acceleration of a particular object which is identical to that produced by a single 5-unit force inclined at arccos 0.6 to the 3-unit force, and arccos 0.8 to the 4-unit force (see figure). The laws of vector addition thus apply to the superposition of forces.



Vector addition of forces.

Conversely, a single force may be considered as equivalent to two or more forces whose vector sum equals the single force. In this way one may select the component of a particular force which may be especially relevant to the physical problem, for example, the component of a railroad car's weight along the direction of the track on a hill.

Statics is the branch of mechanics which treats forces in nonaccelerated systems. Hence, the resultant of all forces is zero, and critical problems are the determination of the component forces on the object or its structural parts in static equilib-

rium. Practical questions concern the ability of structural members to support the forces or tensions. *See* STATICS.

Specially designated forces. If a force is defined for every point of a region and if this so-called vector field is irrotational, the force is designated conservative. Physically, it is shown in the development of mechanics that this property requires that the work done by this force field on a particle traversing a closed path is zero. Mathematically, such a force field can be shown to be expressible as the (conventionally negative) gradient of a scalar function of position V.

$$\mathbf{F} = -\nabla V$$

A force which extracts energy irreversibly from a mechanical system is called dissipative, or nonconservative. Familiar examples are frictional forces, including those of air resistance. Dissipative forces are of great practical interest, although they are often very difficult to take into account precisely in phenomena of mechanics.

The force which must be directed toward the center of curvature to cause a particle to move in a curved path is called centripetal force. For example, if one rotates a stone on the end of a string, the force with which the string pulls radially inward on the stone is centripetal force. The reaction to centripetal force (namely, the force of the stone on the string) is called centrifugal force. *See* CENTRIFUGAL FORCE; CENTRIPETAL FORCE.

Methods of measuring forces. Direct force measurements in mechanics usually reduce ultimately to a weight comparison. Even when the elastic distortion of a spring or of a torsion fiber is used, the calibration of the elastic property will often be through a balance which compares the pull of the spring with a calibrated weight or the torsion of the fiber with a torque arising from a calibrated weight on a moment arm. *See* BALANCE (WEIGHING INSTRUMENT); WEIGHT MEASUREMENT.

In dynamic systems, any means of measuring acceleration—for example, through photographic methods or radar tracking—allows one to calculate the force acting on an object of known mass.

Units of force. In addition to use of the absolute or the gravitational approach, one must contend with two sets of standard objects and lengths, the British and the metric standards. All systems use the second as the unit of time. In the metric absolute system, the units of force are the newton and the dyne. The poundal is the force unit in the British absolute system, whereas the British gravitational system uses the pound. Metric gravitational systems are rarely used. Occasionally one encounters terms such as gram-force or kilogram-force, but no corresponding mass unit has been named. *See* DYNE; NEWTON; POUND; POUNDAL; *see also* UNITS, SYSTEMS OF. [C.E.P.]

Bibliography: M. L. Bullock, Systems of units in mechanics—a summary, *Am. J. Phys.*, 22:291–299, 1954; R. B. Lindsay and H. Margenau, *Foundations of Physics*, rev. ed., 1957.

Force fit

A means for holding mating mechanical parts in fixed position relative to each other. In a force fit of cylindrical parts, the inner member has a greater diameter than the hole of the outer member; that is, the metals of the two parts interfere. *See* ALLOWANCE.

Light drive fits require assembly pressure and produce a more or less permanent assembly of thin sections or long fits. They are suitable for cast iron external members. For a tighter fit, medium or heavy drive fits are used, although such fits may require more than direct mechanical pressure for assembly (*see* SHRINK FIT). In a true force fit, the parts are highly stressed, the interference amounting to 0.002 or 0.003 in for parts with a basic diameter of 1 in [P.H.B.]

Forced oscillation

An oscillation produced in a simple oscillator or equivalent mechanical system by an external periodic driving force. Forced oscillations are to be contrasted with free oscillations which occur when the system is displaced from equilibrium and released. *See* OSCILLATION; *see also* HARMONIC MOTION; MECHANICAL VIBRATION; VIBRATION.

Application of the driving force results at first in two simultaneous oscillations, one at the frequency of the free oscillations, and one at the frequency of the impressed driving force. The former, the so-called transient, eventually decays, whereas the latter builds up to the steady-state value. Small damping is associated with slow decay of the transient and slow establishment of the steady state, whereas large damping is associated with fast decay. *See* DAMPING.

Consider a simple oscillator having mass m, stiffness (ratio of restoring force to displacement) s, and mechanical resistance (the negative of the ratio of resisting force to velocity) R. The undamped (that is, calculated on the basis $R = 0$) natural angular frequency is $\omega_0 = (s/m)^{1/2}$.

If a sinusoidal driving force of amplitude F and angular frequency $\omega = 2\pi f$ is applied, then after the steady state is reached, the velocity amplitude is

$$V = F/|Z|$$

where Z is the complex mechanical impedance and the vertical bars denote absolute value, or magnitude. The quantity Z depends on m, R, and s. *See* IMPEDANCE, MECHANICAL.

The maximum velocity amplitude occurs for $\omega = \omega_0$ and is equal to

$$V_m = F/R$$

The ratio of V to V_m depends on the frequency only through the expression

$$X = \frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \qquad (1)$$

so that the amplitude response at, for examp

twice the resonant frequency is the same as that at *half the resonant frequency* (geometric symmetry), and on the damping only through the so-called quality factor

$$Q = \omega_0 m/R \qquad (2)$$

In these terms the response is

$$V/V_m = (1 + Q^2X^2)^{1/2} \qquad (3)$$

and the phase, ϕ, of the velocity, relative to the driving force, is given by

$$\phi = \arctan\ (-QX) \qquad (4)$$

Equation (3) shows that if Q is large, the response is appreciable only if X is small, so that ω *is not too* different from ω_0. Thus, high Q corresponds to narrow bandwidths. Large bandwidths, centered on ω_0, are obtained with low Qs. According to Eq. (4), the velocity leads, is in phase with, or lags behind the force for frequencies below, on, or above resonance, respectively. *See* RESONANCE (ACOUSTICS AND MECHANICS).

In some cases, the displacement response is the one of interest. The displacement amplitude at zero frequency is $d_0 = F/s$, whereas that at $\omega = \omega_0$, with $x = 0$, is exactly Q times as great. The maximum displacement amplitude occurs at an angular frequency somewhat below ω_0, which is also the natural frequency of the damped system. The difference is small if Q is large. Flat displacement response occurs at frequencies well below resonance.

[M.GR.]

Bibliography: J. P. Den Hartog, *Mechanical Vibrations*, 4th ed., 1956; L. E. Kinsler and A. R. Frey, *Fundamentals of Acoustics*, 1950; N. W. McLachlan, *Theory of Vibrations*, 1951.

Forcipulata

An order of Asteroidea characterized by crossed pedicellariae. Other characters include the stellate outline of the body which is not pentagonal; small, inconspicuous, marginal plates; papulae usually present on both [illegible] of the body; [illegible] one often sur[illegible] lariae; tube-f[illegible] four longitudinal rows along the ambulacral groove. The order includes only four families of starfishes, but the genera are numerous and widely distributed. The Brisingidae and Zoroasteridae are characteristically deep-water forms; the former family includes species with numerous arms (as many as 44). The Asteriidae is the largest family, and includes many well-known predatory sea stars, some of which are economically significant because of the damage they inflict upon oyster beds. *See* ASTEROIDEA.

[H.B.F.]

Forensic chemistry

The application of chemistry to the study of materials or problems, in cases where the findings may be presented as technical evidence in a court of law. It is one branch of the broader field of forensic sciences. Its boundaries are not sharply defined, for it includes topics that are only partly chemical. The forensic chemist conducts scientific examinations of evidence submitted in connection with a criminal investigation or a civil suit. He performs toxicological analyses, studies exhibits from scenes of suspected arson or sabotage, and on occasion he may undertake nonchemical examinations such as a study of markings made by a tool.

Materials submitted for examination. Among the materials most frequently submitted for an examination are the following:

1. Viscera and body fluids to be analyzed for poison; blood and other body fluids for alcoholic content
2. Narcotics in bulk and as minute residues
3. Blood, hair, and seminal stains for identification
4. Liquor, legal and illicit, for alcoholic content
5. Paint, usually as minute chips, or perhaps as a thin smear on an article such as a tool that may have been used to force entry into a building
6. Glass, often as minute fragments found, for example, on clothing of a victim of a hit-and-run accident where perhaps a headlamp was shattered
7. Textile fibers, often as a small tuft such as might be dislodged from the clothing of an assailant; cordage fibers
8. Soil, particularly as small deposits on clothing or shoes
9. Ink on documents, including restoration of obliterated writing by chemical means
10. Materials to be tested to determine their specifications where legal proceedings may ensue.

Among many other items are packing materials from safes, lipstick, tobacco and tobacco ash, explosives and residues therefrom, bullets, and incendiary materials.

The forensic chemist is a versatile analyst and a microscopist; and he must be well versed in a wide variety of materials, industrial and otherwise.

Samples and records. In a criminal investigation it is advisable, if possible, for the forensic chemist to visit the scene of the crime before anything is disturbed. His specialized knowledge is then put to best use, for he takes those articles and samples likely on scientific examination to yield useful information. His field notes on the condition and exact position of each article and sample may later prove invaluable.

If untrained persons are to collect and pack material destined for the laboratory, it is advisable that they first obtain instructions on the proper way of doing so; many forensic laboratories provide printed instructions.

On receiving an article or sample, the forensic chemist enters in a register the date and time of receipt, the name and official position of the submitter, the nature of the material, any identifying marks on it, and the serial number which he assigns. All wrappings and seals are preserved for possible presentation in court.

The forensic chemist maintains full notes in ink of every stage of the investigation, beginning with a description of the wrappings and seals on samples.

Methods and equipment. Examinations usually are for one of two purposes: identification and measurement of the amount of a substance, such as a poison isolated from viscera, or comparison of two samples to determine whether one could have been the source of the other. For example, a comparative laboratory examination may show whether or not a chip of paint could have come from a particular automobile, or a tuft of fibers from a particular garment.

Mass-produced articles tend to lack individual peculiarities. Further, one of the samples for comparison may be too small for a complete examination. Therefore, while the results of a comparative study may be consistent with a sample's having come from a particular source, they will rarely exclude other identical sources. However, each successive independent test that gives identical results on two samples may considerably increase the likelihood of the samples' being identical, because probabilities are multiplicative, not additive.

Judgment is required in choosing tests most likely to provide maximum information, particularly on minute samples. Nondestructive tests are particularly valuable; they include measurements of some physical properties such as density, microscopic studies including the measurement of optical constants of crystals, analysis by x-ray diffraction, and fluorescent x-ray spectrometry.

Major instruments of special value in forensic chemistry include the optical spectrograph (grating or prism), the fluorescent x-ray spectrograph, x-ray diffraction equipment, microscopes, the microtome, and the spectrophotometer for infrared, visible, and ultraviolet absorption spectra.

The optical spectrograph is indispensable. A high quality instrument with relatively high dispersion and good resolving power is advisable, but an instrument with medium dispersion is adequate for materials with simple spectra, such as glass. Accessories for quantitative analysis are needed. Even for the comparison of two samples, densitometry of chosen spectral lines is preferred to visual inspection.

The applicability of the optical spectrograph to the detection of about 70 elements makes possible fairly complete analyses on small samples. Rapid quantitative comparison of two samples is also possible; here, the trace elements present may be the best indicators. However, small samples of some materials may not be representative.

The fluorescent x-ray spectrograph complements the optical spectrograph. It is less sensitive than the latter instrument, and so is not usually chosen for detecting trace elements. However, the sample is usually unaltered by an x-ray analysis, smears of material on exhibits can be analyzed qualitatively without their prior removal, and the accuracy of quantitative analysis by x-ray exceeds that by an optical spectrograph. For both instruments, the preparation of standards can be time-consuming.

X-ray diffraction patterns may confirm the molecular composition of a crystalline substance or a simple mixture; in toxicological analysis they serve to confirm the identity of certain poisons and drugs. Further, diffraction patterns are a means of comparing two samples, such as packing materials from safes for possible identity.

In microscopy, a polarizing microscope is much more useful than a biological microscope. For example, the optical constants of a transparent crystalline substance can be measured with a polarizing microscope. These constants may identify the substance; they may also be a means of comparing two samples for possible identity.

The formation of a minute crystalline precipitate on a microscope slide by reaction of a grain of sample and a reagent is a vivid, sensitive method of qualitative analysis, frequently used in the identification of poisons as well as of inorganic substances.

Fine structure and heterogeneity are usually detectable by microscopic study. For example, fibers can be identified. The reactions of fibers with certain reagents are also diagnostic.

A comparison microscope permits the simultaneous comparison of two samples. Thus transverse sections of two multilayered chips of paint are brought into juxtaposition by the optics, and the color and thickness of each of their various layers may be compared for possible identity.

Metallographic techniques may provide necessary information about metallic samples or be a means of comparing two samples. Microscopic study of polished and etched surfaces by reflected light reveals internal structure, such as grain size. Macroscopic study of etched surfaces reveals not only external structure, but also on a cut section it reveals internal structures; thus, flow lines of a forged alloy may serve to match two severed pieces. Erased markings on metals, such as the serial number on a gun, may be restored by polishing and etching the surface.

Ultraviolet and visible absorption spectrophotometry are used to assay for elements after adding appropriate reagents, and to identify and measure the amounts of certain organic substances, particularly drugs and poisons such as barbiturates. Absorption spectra of the dyestuff in two samples such as lipstick may confirm their identity.

Forensic chemistry is not restricted to instrumental methods. Spot tests are used, especially during the preliminary stages of criminal investigations, as a means of rapidly screening many exhibits; an example is the benzidine test for blood. Classical and new chemical procedures are extensively employed. Microchemical analysis, with its elegant techniques for handling a wide range of sample sizes, is particularly useful in a field where sample size is usually dictated by circumstance.

Of the special techniques employed by forensic chemists, a notable example is the use of density-gradient tubes to obtain comparative density

terns of heterogeneous powders such as soils. In it, a glass tube containing several layers of liquids with graded densities is used. Diffusion erases the original discontinuous boundaries, and the components of a heterogeneous powder distribute themselves in the tube according to their densities, to provide a density pattern useful for comparative purposes.

Toxicological analysis. The forensic chemist performs toxicological analyses to isolate, identify, and measure the amount of a poison. Most samples for analysis are of biological material obtained by a pathologist at autopsy. They include brain, liver, blood, urine, kidney, lung tissue, stomach and contents, and sometimes other material. About 1 pound of each organ is submitted, a record being kept of the weights of all samples. The chemist saves a reasonable amount of each sample for possible presentation in court, in case a second analysis by another chemist is requested.

Some poisons produce characteristic symptoms and pathological changes discernible to the autopsist. Some unconsumed poison may have been found at the scene of death. A list of medicinal preparations used by the deceased may be available. Such evidence may point to a particular poison, which is then sought first by using an abridged analytical procedure. Otherwise the analysis becomes a general unknown, requiring a lengthy systematic search; but cognizance is taken of the high frequency of occurrence of a few poisons, which include carbon monoxide, cyanide, ethyl alcohol, and barbiturates.

For analytical purposes, poisons are usually divided into five groups, based on methods used to separate them from biological material.

Group I, volatile poisons, separable from tissue by steam distillation. Examples are alcohols, nicotine, hydrocyanic acid, and solvents such as carbon tetrachloride.

Group II, gaseous poisons which have been inhaled, such as carbon monoxide.

Group III, organic poisons not volatile with steam, separable from tissue by extraction with ethyl alcohol. This large group includes the alkaloids, the ever-increasing number of synthetic drugs such as barbiturates, and other substances such as ptomaines, glycosides, and some pesticides.

Group IV, inorganic poisons, comprising all poisonous metals and nonmetals and their poisonous compounds.

Group V, miscellaneous poisons, namely those substances not isolated by the methods used for the other four groups.

Identification of a poison in Group I is usually by testing aliquots of the steam distillate. On occasion, physical constants are measured on the poison purified by fractionation of that distillate. Gas chromatography is finding an important place in the identification and measurement of volatile poisons.

Chemical tests for intoxication. Standard chemical methods have long existed for the analysis in the laboratory of blood, urine, and saliva. The use of breath for routine analysis is now becoming widespread. Breath analysis is convenient and simple by means of commercially available instruments; comparison of results with those obtained by standard laboratory methods has indicated a satisfactory degree of accuracy. The several available instruments differ in some details and in ease of operation; not all of them can be used on an un-

holic content of this air, the fixed ratio permits the alcoholic content of the blood to be calculated. *See* ANALYTICAL CHEMISTRY; FORENSIC MEDICINE; PHARMACOLOGY; TOXICOLOGY. [C.C.B.C.]

Bibliography: P. L. Kirk, *Crime Investigation*, 1953; L. C. Nickolls, *The Scientific Investigation of Crime*, 1956; C. E. O'Hara and J. W. Osterburg, *An Introduction to Criminalistics*, 1949.

Forensic medicine

The application of medical evidence or of medical opinion for purposes of law, either civil or criminal. Any of the numerous branches of medicine may participate in pretrial settlement or court proceedings. The expert medical witness responds either by agreement with an attorney or in answer to a subpoena or summons. It has been estimated that in 80% of appellate cases alone, medical evidence is an integral part of the pleading. Medical specialties most frequently contributing evidence in legal disputes are forensic pathology, forensic psychiatry, orthopedic surgery, general surgery, general practice, and dentistry. In medical malpractice actions, the entire plaintiff and defense cases are based on medical evidence. In industrial compensation hearings, the claimant usually bases his case on the theory that stress, frank injury, or exposure to noxious substances in the course of employment caused disability or death either directly or by the aggravation of preexisting disease. Similar claims are brought before governmental pension boards and Veterans Administration appeal boards.

Injury and disease relations. Probably the most complex problem in forensic medicine is to establish whether in a given instance there was an interrelation between injury and disease, either directly or by the aggravation of a preexisting disease by injury. The classic example concerns injury and cancer. The appearance of cancer in an area of the body which was previously the site of external violence makes it difficult to convince the laity that violence was not the cause of the disease, yet competent medical authorities agree that single injury does not result in malignancy. The second phase of this interrelation is concerned with the question of whether preexisting cancer can be made worse (for example, faster-spreading) by an injury to the general area of the body where the disease exists. This question is debatable. Cancer near the surface of the body can certainly be reached by an external force and thus con-

ceivably be caused to spread, or in other words, to be aggravated. However, an already disseminated malignant disease such as leukemia cannot reasonably be claimed to spread as the result of injury. In the problem of injury and cancer, as in all other disease-injury relationships, a valid opinion can usually be obtained from the medical specialist who concentrates his practice on the matter at hand It must not be inferred, however, that all disease-injury problems are complicated. The orthopedic or general surgeon or (in fatal cases) the forensic pathologist will quickly recognize the direct connection between clots of blood in the pulmonary arteries and the fairly recent leg fracture. Occasionally injury to a leg causes the damaged veins to form thromboses that propagate upward, break free, and float to the lungs where they impede the flow of blood. The resulting pulmonary embolism damages the lung or causes death. *See* CIRCULATION DISORDERS; EMBOLISM; LUNG DISORDERS; THROMBOSIS.

Cardiac disease. Cardiac disease is another constant problem in the disease injury relationship. It is probable that unaccustomed stress placed upon the heart of an elderly man with inadequate circulation to the heart muscle through narrowed coronary arteries can precipitate his death; for example, after shoveling snow for hours he may drop dead. Cardiac damage or death is most likely in those with narrowed coronary arteries who, after leading a sedentary life, suddenly experience severe exertion or extremely grave emotion. On the other hand, judicial or quasi-judicial bodies have granted monetary benefits on the flimsiest of grounds, such as the case of the motor vehicle inspector who died while sitting in an armchair listening to the questioning of an informant by a third individual. It was argued that the emotional stress existed because the investigator was acquainted with the woman killed in the car, and industrial compensation was quite unjustifiably awarded. *See* CIRCULATION DISORDERS; HEART DISORDERS.

Medical experts. Government usually establishes a constitutional or statutory office to investigate deaths known or thought to be of public concern. The medical expert designated to assume the responsibility for these deaths (about 25% of all deaths) is usually termed coroner or medical examiner. Whatever his designation, he will provide more medical evidence for the courts and police than any other individual in his community, be it state, county, or municipality. He should be a graduate of an accepted medical school with minimal post-graduate training consisting of 1 year of clinical internship, 2 years of hospital pathology, and 1 year of training and 1 of practice in the specialized procedures of forensic pathology. Such training and experience qualifies the physician for examination and certification as a specialist by the American Board of Pathology. Such a medical official should then be qualified, when indicated, to perform six functions: exonerate the innocent; recognize violence and the type of violence; acquire, interpret, and present in court nonpartisan and intelligible medical evidence; recognize fatal industrial health and safety hazards; recognize fatal infectious disease; and utilize the material for teaching law and medical students, physicians and lawyers, and police.

Procedures. Approximately one-half of the deaths deserving public scrutiny are found to be due to natural causes. The remaining fraction includes homicide, suicide, accidental death including traffic fatalities, industrial deaths, and death due to infectious disease. Special problems rarely faced in other branches of medicine include identification of decomposed bodies and parts of bodies and estimation of the time of death and time of injury. Anthropological measurements are used to determine stature, age, and sex of skeletal remains The reconstruction of damaged and decayed fingertips to record fingerprints is a special technique. The degree of tissue response to injury may indicate the duration of the interval from injury to death, the decrease in internal body temperature or the extent of postmortem degeneration may provide an approximation of the interval from death to discovery. In addition to their obvious criminal significance, the special procedures for identification and time of death are important for insurance recovery and the settlement of estates.

One-half of all deaths of public concern require some chemical procedure either to establish or to explain the cause and manner of death. A blood- or tissue-alcohol level consistent with gross incoordination and mental confusion does much to explain the death of a pedestrian who stumbles against the side of a passing bus. A massive dose of a barbiturate in the stomach suggests suicide. Cyanide, strychnine, and nicotine may be [illegible] the

ica-

to

which employees are exposed creates an unending demand for experimental study of their potential harmful effects. The fluorescent-lamp industry unwittingly exposed its workers to beryllium compounds so hazardous that months later secretaries and laundry workers were disabled or died.

Current legal status. The current status of official investigative systems for the determination of the cause and manner of death is extremely varied. In some states a death without obvious signs of violence may not be subject to autopsy, yet widespread damage to viscera frequently occurs without external signs. About 50,000,000 people in the United States are served by forensic pathologists with adequate training operating under permissive laws. [R.FO.]

Forest and forestry

Forestry involves the management of forest lands for wood, forage, water, wildlife, and recreation. A forest is much more than an assemblage of trees. It is rather a community—technically an [illegible] consisting of plants and animals and

ment in which they live (*see* FOREST ECOLOGY). Trees are the dominant form of vegetation, but shrubs, herbs, mosses, fungi, insects, reptiles, birds, mammals, soil, water, and air are essential parts of the community, each of which reacts on all the others (*see* FOREST SOIL). Therefore, the first prerequisite for forestry is knowledge of the characteristics of the living members of the community, of the environment in which they live, and of the interrelations among organisms and environments.

Management. The many goods and services afforded by the forest are considered, in light of their relative values, and a decision made as to which should be favored in the management of the forest under consideration (*see* SILVICULTURE). Wood for direct use or for manufacture into innumerable materials is the major product of forests, and has been the chief concern of forest management techniques (Fig. 1). However, there is growing recognition of the value of other tree products, such as naval stores, maple syrup, quinine, paper, and rubber (*see* WOOD CHEMICALS; WOOD FIBER PRODUCTS). Forage both for domestic livestock and for wildlife is often important. In addition to these tangible products, numerous services are rendered by the forest.

Forests increase the water-holding capacity of the underlying soil and facilitate the entrance of water into it. They thus reduce the surface runoff of water and tend to prevent erosion and to regulate the flow of springs and streams (*see* CONSERVATION OF RESOURCES). By decreasing wind movement and reducing evaporation from the surrounding area, forests temper the local climate and increase the production of crops in their lee (*see* AGRICULTURE). Forests also furnish some of the most attractive opportunities for hunting, fishing, picnicking, camping, and other recreational activities (*see* FOREST CONSERVATION).

Management may be centered on any one or a combination of these various goods and services. The usual criterion—which pays the better?—used in comparing the desirability of different courses of action is difficult to apply here, even when costs



Fig. 1. Loading Douglas-fir log onto logging truck, Olympic National Forest, Washington. (U.S. Forest Service)



Fig. 2. Openings in many forests provide forage for the grazing of livestock, Challis National Forest, Idaho (U.S. Forest Service)

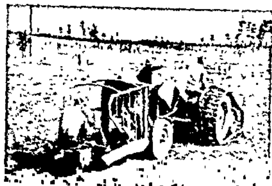
and returns can be expressed in dollars and cents, as they may be with timber products. It becomes much more difficult when the returns cannot be so expressed, and when they accrue to someone other than the owner, as is commonly the case with watershed protection and with recreation. When mathematics fails, the forest manager must rely to a considerable degree upon his own judgment.

Following adoption of a basic policy for the handling of a specific forest property, the next step is to place it under actual management. It is appropriate to mention some of the major considerations that control forest policies and practices.

Multiple use. The variety of goods and services obtainable from the forest gives special importance to the principle of multiple use. In essence, this principle requires the use of the different parts of a forest for the purpose or combination of purposes for which each is best suited. Timber may be grown for a single product or, preferably, for the integrated utilization of several products such as sawlogs, veneer logs, piling, posts, and fuelwood (*see* LUMBER MANUFACTURE). Often there are areas in the forest in which management for the simultaneous production of timber, forage, wildlife, water, and recreation is both feasible and desirable. Multiple use, however, has definite limitations. Some uses, under certain conditions, are incompatible, and maximum production of all kinds of goods and services on the same area is impossible.

Modern forestry, although chiefly concerned with timber management, often includes some aspects of range, wildlife, watershed, and recreation management. These activities may be conducted in the forest itself or in forest openings, which for administrative reasons must be handled as part of the forest property (Fig. 2). Because of the close relation between the production and the utilization of wood, the forester is also concerned with the properties of wood and the processes by which it is manufactured into thousands of products. Although wood technology is not strictly a part of forestry, the two are so closely related that they are commonly handled together in education and research (*see* WOOD PHYSICS).

Sustained yield. The basic principle of sustained yield involves the continuous production of forest goods and services. In timber management,



ing reforestation. DeSoto National Forest, Mississippi. (U.S. Forest Service)



Fig. 4. Light thinnings in dense stands forestall mortality and concentrate growth on the better trees. Timber stand improvement, McCleary Experimental Forest, Washington. (U.S. Forest Service)

not only must reforestation, either by natural or artificial means, follow cutting, but the forest property as a whole must be organized so that approximately the same amount of wood may be harvested year after year (Fig. 3). Growth and drain must balance at a sufficiently high level to meet the current demand (see FOREST SEEDING AND PLANTING). The same principle that provides for continuous wood production without reduction in quantity or impairment in quality also applies to other products such as wildlife, water, and recreation. As population increases and standards of living rise, pressure on the forest pushes to higher and higher levels the sustained yield which will be adequate to meet mounting needs. The higher the level and the larger the number of products and services involved, the more difficult the task of management becomes (Fig. 4).

Government participation. Forests play so vital a role in so many ways in the well-being of a nation that governmental action to assure their effective management has long been recognized as desirable, virtually throughout the world. That action may take the form of public ownership, public regulation, or public cooperation.

In the United States, governmental activity has been chiefly in the fields of ownership and cooperation, with little attempt at regulation. Since 1891, the Federal government has reserved large areas of land in the public domain, and since 1911, it has pu[illegible]
areas [illegible]
out co[illegible]
000 acres. The states (notably Pennsylvania, Michigan, Minnesota, and Washington) owned an additional 19,000,000 acres of commercial forest land, and counties and other local governmental units 8,000,000 acres. While the area in these different forms of ownership is subject to continual change, there is little doubt that the policy of public ownership on a substantial scale is a permanent one.

Cooperation with private owners is extended by both state and Federal governments in a wide variety of ways (see FORESTRY, FARM). Most prominent is the protection of forests from fire (Fig. 5), an activity handled largely by the states, with the assistance of grants-in-aid from the Federal government (see FOREST FIRE CONTROL). Other cooperative activities include participation in protection of forests from insects and diseases (see TREE), sale of planting stock at cost of production, and educational and service assistance to private owners in the production, harvesting, and marketing of forest crops (Fig 6) Some states have special legislation relating to the taxation of forest lands which is intended to encourage their improved management. A few states exercise some control over



Fig. 5. Lookout towers from which fires are spotted and reported constitute an important element in the essential task of fire control. Fire detection tower in the Olympic National Forest, Washington. (U.S. Forest Service)



Fig. 6. Spraying by airplane is a modern device for the control of forest insect epidemics. Boise Nat' Forest, Idaho. (U.S. Forest Service)

the management of private lands by establishing minimum standards of forest practice.

As a managerial activity dealing with many products and services of major value, forestry makes full use of the biological and physical sciences, of engineering, and of the social sciences. It requires technical skill of a high order in the practical application of scientific knowledge with due regard to economic and social considerations. See separate articles on important forest trees listed under their common names *See also* DENDROLOGY; FOREST MAPPING; FOREST MENSURATION; FOREST TREE NURSERY MANAGEMENT; WOOD FINISHING; WOOD PRESERVATION. [S.T.D.]

Bibliography: S. W. Allen, *An Introduction to American Forestry*, 2d ed., 1950; F. S. Baker, *Principles of Silviculture*, 1950; H. P Brown, A. J. Panshin, and C. C. Forsaith, *Textbook of Wood Technology*, 2 vols., 1949, 1952; A. H Graves, *Illustrated Guide to Trees and Shrubs*, rev. ed., 1956; W. M. Harlow and E. S. Harrar, *Textbook of Dendrology*, 4th ed., 1958; R. C. Hawley and D. M. Smith, *The Practice of Silviculture*, 1954; H L. Shirley, *Forestry and Its Career Opportunities*, 1952; A. E. Wackerman, *Harvesting Timber Crops*, 1949; *Trees*, USDA Yearbook of Agriculture, 1949

Forest conservation

The renewal, management. and use of forests in such a way as to produce continuously goods and services of economic and social value. This involves proper cutting practices and protection from fire, insects, diseases, and excessive grazing by domestic livestock or wild animals (*see* FOREST FIRE CONTROL; PLANT DISEASE; RANGE LAND CONSERVATION; SILVICULTURE).

Forests are unequally distributed throughout the world. Some nations have abundant forest lands, while others must depend on wood imports. Over 60% of the softwood timber upon which the world depends for construction material is in North America and Europe, areas containing only one-third of the world's population. Densely populated China and India possess less than 3% of the world's forests, whereas sparsely populated South America embraces almost 30% (Fig. 1).

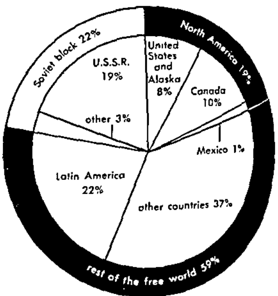


Fig. 1. World forest area, 9,600,000,000 acres North America has one-fifth of the world forest area

Major world forest formations. Forests and other vegetation may be grouped into broad classes or formations These are expressions primarily of the climate prevailing in the areas where the vegetation grows. A good example is the tundra, the vegetation north of the forested region in the Northern Hemisphere, which consists of low shrubs, mosses, sedges, and grasses. South of the tundra is the coniferous forest formation (*see* CONIFERALES; DENDROLOGY). The forests of the world may be classified into three major formations, namely, boreal conifers, temperate hardwoods, and tropical hardwoods. See separate articles on important forest tree species.

Boreal conifer forests. A cool temperate climate determines the natural occurrence of most coniferous species; hence they are found in a broad belt encircling the earth immediately south of the treeless tundras of the far north. In the northern countries the term conifer refers mainly to spruces and firs. The southern part of the boreal conifer forests also contains pines, larches, and hemlocks in addition to the spruces and firs which overlap into the temperate hardwood forests.

Temperate hardwoods. Deciduous forests are characteristic of temperate regions having uniform seasonal distribution of rainfall. In Europe, North America, and the U.S.S.R. they consist of many species of birch, beech, maple, ash, oak, elm, and basswood.

Tropical hardwoods. These forests occur throughout the tropics in lowland areas, especially along the larger streams in India, Burma, Philippines, the coast of Australia, and the streams of South America. The most striking feature of most tropical forests is the great number of species.

United States forest regions. The forests of the United States occur in five broad regions. Three of these are east of the prairies: the northern forest, the central hardwood forest, and the southern forest. The Rocky Mountain and Pacific forests are west of the Great Plains, extending from Canada south to Mexico (Fig. 2).

The northern forest region covers almost all of New England, New York, Minnesota, Wisconsin, and Michigan, and it extends southward in the Adirondack Mountains and the Appalachian Highland as far as northern Georgia. White and black spruce, balsam fir, eastern white pine, red pine, jack pine, and hemlock are the principal conifers found in this region. Sugar maple, beech, and yellow birch are the most common hardwoods associated with the conifers.

The central hardwood forest, principally of deciduous trees, is found in 19 states in the central eastern part of the United States. This forest consists of a wide variety of species including beech,

ROCKY MOUNTAIN
NORTHERN
TROPICAL

Fig 2 The natural forest regions of the United States (U S Forest Service)

sugar maple, American elm, and many species of oak and hickory. Other species found in this region are wild black cherry, cottonwood, tulip tree, shag-bark hickory, white ash, black walnut, red maple, silver maple, and cucumber tree.

The southern forest extends from Virginia to eastern Oklahoma and Texas and thence to the Gulf of Mexico. This vast forest is characterized by four important pines: shortleaf, longleaf, lob-lolly, and slash. It also encompasses cypress swamps and bottomland hardwoods.

The Rocky Mountain forests extend from Canada to Mexico and are primarily coniferous, with pon-derosa pine the most common tree. At the higher elevations, Colorado blue spruce, Engelmann spruce, and Douglas-fir are dominant. Magnificent forests of western white pine, western larch, west-ern hemlock, and ponderosa pine also flourish in various combinations at lower elevations.

The Pacific Coast forest contains the heaviest forest stands in America and probably in the world. Redwood, giant sequoia, Douglas-fir, and sugar pine are important forest trees on the west slopes of the mountains. The magnificent redwood forests commonly contain trees 175–225 ft in height and 3-10 ft in diameter, but many attain considerably larger sizes.

Forest acreage and timber volume About 664,-000,000 acres are primarily useful as forest land. Approximately 175,000,000 acres of this is noncom-mercial, including forest land in rugged, mountain-ous country or located where climate is unfavorable or the soil too shallow for other crops. This leaves 489,000,000 acres of better or more accessible tim-ber-growing land, or about 3 acres for every person in the United States (Fig. 3).

Forests cover one-fourth of the land area in the United States. Timber volume amounts to slightly more than 500,000,000,000 cu ft. Pine, fir, and other softwoods constitute 80% of the forest sawtimber. Three states—Oregon, California, and Washington—have about half of the nation's sawtimber vol-ume, while the industrialized eastern half of the nation has three-fourths of the forest acreage.

Public forests Of the total forest land, 27%, or 130,000,000 acres, is in public ownership. The

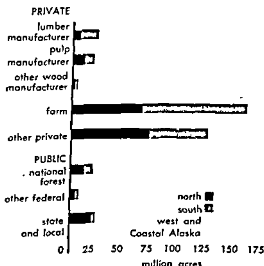


Fig. 3. Commercial forest land ownership.

largest acreage of public timber is in the national forests—17% of the total. The aggregate area of state forests is now more than 15,000,000 acres.

Privately owned forest land thus comprises 73%, or 359,000,000 acres, of the total. Large industrial forests amount to 13%, or 63,000,000 acres. These are owned mostly by lumber and paper companies, of which there are 23,000. Farm woodlands and small acreages contain 60% or 296,000,000 acres in 4,500,000 ownerships. Small forests of less than 100 acres each, most of them farm and nonfarm woodlands, make up 86% of the total number of private ownerships.

Supply, demand, and growth. The forests of the United States are now growing annually at the rate of 13,400,000,000 cu ft, compared with an estimated yearly consumption of 12,000,000,000 cu ft of timber products. Not all of the current growth is in economically usable form. The average American uses about 80 cu ft of wood per year.

Lumber drain is estimated at 48,800,000,000 board-feet of sawtimber per year, whereas growth is 47,400,000,000 board-feet. Most eastern species now have favorable growth-cut ratios.

Pulpwood situation. Since about 1920 there has been a steady increase in annual consumption of wood pulp in the United States. The use of wood pulp for the manufacture of paper, paperboard, derived products of wood pulp such as rayon and cellophane, and wood-fiber products including hard board and roofing felt, has more than quadrupled, having increased from 8,100,000 cords in 1920 to 37,800,000 cords in 1953 (Fig. 4).

Forest uses. Although high standards of living require the abundant use of wood, forests include other resources of great economic and social significance that in some regions transcend their value as wood producers.

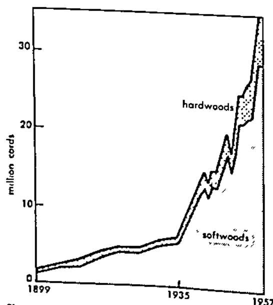


Fig. 4. Pulpwood production trends in the United States from 1899 to 1957. (Bureau of the Census, U.S. Department of Commerce, and U.S. Forest Service)

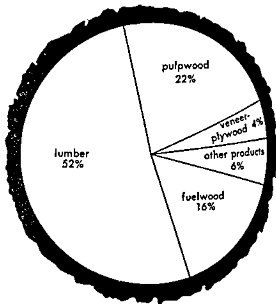


Fig. 5. Lumber constitutes over half the total wood use. (U.S. Forest Service)

Multiple use. Forests conserve and regulate inland waters essential for irrigation, navigation, power, aquatic life, and domestic use; they form protecting cover that helps greatly to control soil erosion; they are the natural home of much wildlife; and, through their increasing importance in outdoor recreation, they promote the health and spiritual well-being of the people. Fortunately, most of these values can often be derived simultaneously with wood products from the same land. In general, it can be said that if there is sound timber management, the other forest values and uses will develop almost automatically.

Forest products. Since colonial times, wood has performed a vital role in the development of the United States. Wood is an extremely adaptable material. No other material provides fuel, fiber, food, and chemical derivatives while also serving a wide variety of structural uses. A few of the principal products of the forest harvest are shown in Fig. 5.

The volume of timber products consumed in 1952, expressed in terms of the cubic-foot volume of logs and bolts, amounted to 12,300,000,000 cu ft, or 78 cu ft per capita.

The turpentine or gum naval-stores industry is concentrated in the slash and longleaf pine belts of Georgia, Florida, and Alabama (see NAVAL STORES). The United States in normal times supplies the world with one-half its needs for turpentine and rosin. In 1947–1948 the industry produced 294,028 barrels of turpentine and 829,128 drums of rosin, bringing a total return to the South of $39,000,000.

The maple sugar and sirup industry is limited chiefly to the following states in order of production: Vermont, New York, Pennsylvania, Michigan, Ohio, New Hampshire, Wisconsin, Massachusetts, Maine, Maryland, and Minnesota. According to the United States census, the value of maple sugar and sirup in 1954 was $5,664,899.

Christmas tree farming is becoming an increasingly popular enterprise in many parts of the United States, chiefly in the states bordering Canada, except North Dakota. However, since about 1948, many Christmas tree plantations have been established in the east, central, and midwestern states. A survey by the United States Department of Agriculture showed that more than 225,000 acres are planted for Christmas tree production, 158,000 acres being planted by farmers. About 38,000,000 trees are harvested annually, or enough to supply 9 out of 10 American homes. Of this production approximately 10% are cut from public lands. The remainder are cut from private lands with an increasing number being grown as managed plantations on farms. Assuming an average wholesale value of $1.50 per tree, the annual income from Christmas trees is about $57,000,000. In 1955 the United States also imported 12,500,000 Christmas trees from Canada.

Better and more efficient wood use during recent years has resulted from research in forest products utilization. Pulping processes have been improved to make possible a much wider range of species, and effective methods of modifying wood properties have been found. The development of preservatives and treatment methods has greatly extended the life of wood in outdoor services such as telegraph poles, crossties, and fence posts. Fundamental studies in wood properties, wood structure, wood chemistry, and wood physics have resulted in many utilization advances.

Forest land problems. Prevention and control of destructive agencies such as fire, insects, and diseases, and the use of proper cutting practices, constitute major problems. During the early forest history of the United States, millions of acres of forest land were burned repeatedly. Waste from poor management and utilization in the woods is shown by the fact that one-fourth of all timber cut is not utilized. Furthermore, insects, disease, fire, and other enemies are now killing about one-fourth of our annual net growth. As a result of past mismanagement and the effects of destructive influences, one-tenth of America's standing timber is made up of trees that are rotten or crooked (Fig. 6).

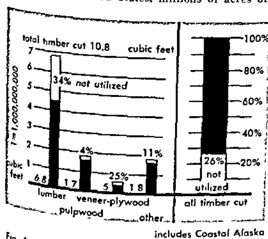


Fig. 6 One-fourth of timber cut is not utilized.

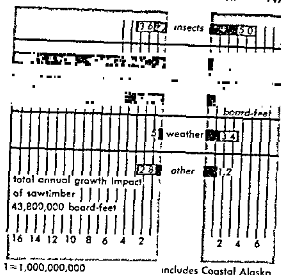


Fig. 7. Factors responsible for the destruction of timber.

In the mill, wasteful utilization is shown by the fact that two-fifths of all plant residues in the United States are unused.

Lands needing reforestation total about 50,000,000 acres, an area equivalent to 10% of the nation's timberlands. Planting on additional millions of acres would also help to improve stocking. The need for education, especially for managers of small tracts, is of great importance because small private forests representing 60% of the timberlands are low in productivity.

Organization for protection from fire and other destructive agents has resulted in a downward trend in the area of land burned by forest fires. Today 88% of the forest area is protected from fires. More intensive efforts will, however, be required to reduce damage from insects and disease (Fig. 7).

Forest conservation goals. Forest conservation measures should have as their goal the intelligent, economical use of all the goods and services a forest can provide on a sustained yield basis. For

Considerable progress has been made in methods of controlling insect pests. Airplane dusting and spraying with DDT have been successful in combating a number of leaf-eating insects (*see* INSECTI-

CIDE). Although increased research and better organization of detection and control are needed, the best possible insurance against insect and disease losses is intensive forest management.

Reforestation through systematic planting has been greatly accelerated during the past decade by the development of mechanical tree planting machines which are well adapted to level or gently rolling lands. Seeding from the air has been successful under certain conditions.

Forest conservation measures. In recent years public and private interests have assumed greater responsibility for forest conservation This has resulted in both public cooperative programs and in industry programs Public cooperative programs consist of federal-state cooperation in fire prevention and suppression, forest management assistance, tree distribution, extension forestry (education), conservation payments, forest insect and disease control, and forest research. The most significant private development has been the Tree Farm program sponsored by the American Forest Products Industries, Inc.. which has as its aim the promotion and development of sound economic forest practices, and the development on a national scale of better understanding and cooperation among private forest enterprises.

Zoning has been in effect on a number of national forests since the 1920s. To date, 79 wilderness areas have been designated by administrative orders, yet a law directing the Secretary of Agriculture to establish such areas and protect them against misuse is considered essential by many.

Public regulation of cutting on private lands has been recommended from time to time as an essential step in continued forest production However, many people feel that the way to improve the situation is rather to provide increased assistance through education, technical guidance, tree distribution, conservation payments, protection, and research. *See* FISHERIES CONSERVATION; MINERAL RESOURCES CONSERVATION; SOIL CONSERVATION; WATER CONSERVATION; WILDLIFE CONSERVATION; *see also* FOREST AND FORESTRY; LAND USE PLANNING. [O.D.D.]

Forest ecology

The science that deals with the relationship of forest trees to their environment, to one another, and to other plants and animals in the forest. The term silvics is synonymous, but is occasionally also used to include forest tree genetics and forest tree physiology as well as forest ecology. All these areas include the fundamental biological knowledge upon which rests the manipulation of the forest for man's own purposes (silviculture).

Forest ecology is commonly subdivided into forest autecology, the study of trees as individuals in relation to their environment, and forest synecology, the study of the forest as a community.

The forest environment or site consists of the physical environment surrounding the aerial portions of the tree (climatic factors) and that surrounding the subterranean portion (edaphic or soil factors). Site is subject to continual change. Atmospheric conditions within the forest fluctuate diurnally, seasonally, annually, and over a period of years; and they are changed by the developing forest itself. The forest soil owes its original nature to the parent geological material and to the weathering processes that formed the soil. It, too, is affected by the climate, the vegetation and animal life it supports, and the passage of time. External influences, particularly fire, grazing and browsing animals, and man, affect markedly the nature of the forest site and its capacity to support tree growth.

Site evaluation. Forest site quality, the capacity of an area to produce forests, is best measured by the recorded growth of trees under forest conditions. Permanent sample plots, 0.1–0.5 hectares (0.246–1.23 acres) in size, are measured every 5 to 10 years for many decades to provide the most accurate measure of site for the conditions they sample. Such plots, however, are relatively few. The height of trees grown free from overhead shade to a specified age (usually 50 or 100 years) provides the best general index of site. In the boreal forest, particularly in northern Europe, the presence of certain "plant indicators" has shown an excellent correlation with site quality.

In situations where forest trees and associated plants cannot be used as an index to site quality, reliance must be placed on soil properties (*see* FOREST SOIL). In a given climatic zone, one or more soil properties may be found to exhibit strong correlation with over-all site quality. Among the properties that have been successfully used as indices in given situations are the parent geologic material, direction and degree of slope, relative topographic position, depth of soil, size of the soil particles, water-holding capacity of the soil, and depth to soil water table. Among climatic factors, the amount and distribution of annual precipitation have often been related to site.

Influence of site upon life history. From seed to old age, forest trees are greatly influenced by the site factor. Vigorous trees growing under favorable site conditions produce large numbers of seeds However, very few seeds, usually far less than 1%, actually grow into trees. Many seeds are destroyed by insects, fungi, rodents, and adverse weather. Only those seeds that come to rest under suitable conditions of moisture and temperature will germinate. Most of the resulting seedlings will be killed by soil fungi ("damping-off" species), solar heating at the ground line, animals, and drought. The most favorable sites for successful forest regeneration usually have moist but well-drained and exposed mineral soil with sufficient dead material on the surface to provide some shade. Regeneration will usually fail where living plants already fully occupy the surface or the surface root zone.

As the seedlings develop, site plays an important part in determining the growth rate of the different species and the outcome of the competition that reduces the typical initial reproduction stand

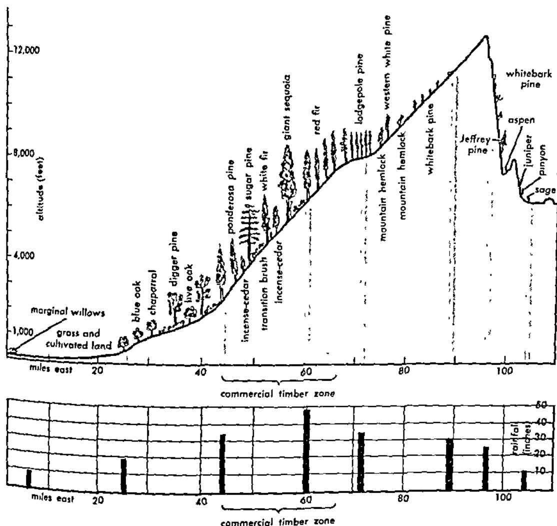


Fig. 1. Profile of the central Sierra Nevada showing altitudinal limits of the principal forest types (USDA)

from several thousand seedlings to a few hundred mature trees per hectare (2.46 acres). *See* ECOLOGIC INTERACTIONS. Some trees, including most of the pines and oaks, will grow well under a wide variety of site conditions, while others, such as black walnut, sycamore, and white ash, can thrive only under specific site conditions. Consequently, the former groups are found on many sites, whereas the latter are less widely distributed.

Different groupings of forest tree species (forest types or associations) thus tend to develop on different sites (*see* DENDROLOGY). In the northeastern United States, for example, a spruce-fir type is characteristic of higher elevations and of cold, wet bogs. Pine types predominate on drier sites, whether sandy or rocky, unless eliminated by logging or fire. Hardwoods occupy the moister lowland soils, the oaks frequently predominating on the somewhat warmer and drier situations, and the birches, beech, and maples on the somewhat cooler and moister. In the western United States, spruces and firs are again found at the higher elevations, with many pines and woodland oaks being found in successively lower belts (Fig. 1).

Influence of forest on site. As the forest becomes established and develops, the site itself is greatly changed. Within the forest, daytime temperatures are lower and nighttime temperatures higher than in the open, thus resulting in more uniform conditions. Air temperature in the forest during the growing season often averages somewhat cooler than in the open. An abrupt forest front may occasionally cause a slight increase (about 1%) in local precipitation by causing moisture-laden air to rise and become cooled. A far more important effect of trees on precipitation is observed in regions where fog or heavy mist occurs. Tiny droplets of atmospheric moisture are deposited on the foliage and on the branches and, when the accumulation is sufficient, water drops to the ground, as "fog drip." Less of the rainfall and snowfall reaches the ground, however, as the leafy crowns of the trees can intercept up to several centimeters of gently falling rain or snow.

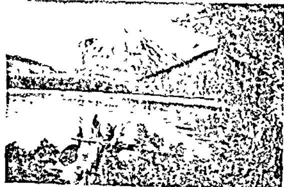


Fig. 2. Much urban and agricultural water flows from forested lands (*U.S. Forest Service*)

On the forest floor, the accumulating layer of leaves, twigs, and other litter attracts a characteristic grouping of plants and animals that live on such decaying organic matter and on each other. The organic matter is gradually incorporated into the top layer of mineral soil, creating a soil horizon rich in nutrients and humus. Water, leaching downward through the soil, tends to deplete the topsoil of soluble materials, depositing these in a lower soil horizon. Some of these materials are taken up by deeply penetrating tree roots, utilized in the development of leaves, and later returned to the surface soil in the annual leaf fall.

Under spruces and firs in cool moist climates, the forest soil is characterized by a whitish leached layer (podzol) virtually devoid of nutrients and located under the layer of unincorporated humus. Under most forest species in the warmer parts of the temperate zone, this underlying zone of leaching is less pronounced. Under the tropical rain forest, the soil is frequently leached of all but iron and aluminum minerals as a result of the wet and warm climate. Here, soil fertility is maintained largely by the rotation of soluble nutrients through the tree roots to the tree crowns and back to the soil as litter. After land clearing for agriculture, many tropical rain-forest soils become infertile in a few months as a result of accelerated leaching and the destruction of the tree root activity that formerly brought the nutrients back to the surface

Forest hydrology. The influences of the forest on the site result in a characteristic pattern of streamflow differing from that on unforested land in that peak streamflow is lower, and the flow after storms is prolonged. Since much urban and agricultural water is produced by forested lands, particularly those in mountains and at other higher elevations, the management of forests for water production is attaining great importance (Fig. 2).

Forests generally use more water than comparable grassland or open land because tree crowns are effective evaporating systems, and because tree roots usually penetrate the soil deeper than those of smaller plants and can thus reach a greater amount of soil moisture. Such transpiration (water-vapor loss from tree crowns) frequently will utilize from one-quarter to one-third the annual precipitation (*see* PLANT, WATER RELATIONS OF). In contrast, grass cover under similar conditions may utilize as little as half as much, primarily because of the shallower depth to which most grass roots penetrate.

On the other hand, forest soils are generally much less compact than comparable grassland or open soils, and can quickly absorb and retain far greater quantities of water. Forest soils can ordinarily absorb all available water unless the soil surface has been compacted by grazing animals or man's activities. In contrast, surface runoff under nonforest conditions may often result in severe soil erosion and in flash floods downstream.

The present tendency in forest management for water production purposes, therefore, is to avoid clear-cutting because of the accelerated erosion and flooding that may result, and instead to thin out the forest so as to maintain favorable forest soil conditions and reduce the number of trees using water. Also, attempts are being made to manage water use by trees whose roots reach the soil water table.

Forest succession. Since the developing forest changes the forest site, it follows that the changed site may itself be more favorable to a new group of tree species other than those currently occupying it. As a result, successive forest associations occupy the site in the absence of fire, logging, windstorm, or other outside disturbance. Thus, in the northeastern United States, pioneer associations characterized by aspen, cherries, birches, and pines occupy open forest sites. These in time will be replaced by intermediate associations containing oaks, maples, ash, and other species, until in several hundred years, the climax association is attained. This latter is the forest type that is capable of maintaining itself without major change, in the absence of disturbance. Species such as hemlock, sugar maple, and beech have the capacity of reproducing under their own cover and thus qualify as members of the climax type.

In much of the world, the influence of fire, grazing animals, and man's activities is so great that the climax is seldom reached. In much of the southeastern United States, for instance, pine types pre-



Fig 3. An abandoned farm reverting to a mixed hardwood-pine forest. (*Farm Security Administration*)

dominate and constitute the most important commercial forests. Yet these owe their existence to frequent burning and logging and will be replaced by a mixed hardwood climax on most sites unless fire or similar disturbances are introduced from time to time (Fig. 3). Similarly, in the western United States, the extensive pure stands of Douglas-fir owe their existence to periodic clear-felling by wind, fire, or man; they would be replaced by less valuable, mixed coniferous types if all disturbances were eliminated. In present-day silviculture, the professional forester utilizes logging, fire, chemicals, and mechanical treatment to maintain the forest on a given site at the stage of forest succession most valuable to man. See ECOLOGY; *see also* FOREST AND FORESTRY; SILVICULTURE. [S.H.S.]

Bibliography: See FOREST AND FORESTRY.

Forest fire control

In the decade 1948–1957, forest fires in the United States totaled 1,632,000, burning over 109,000,000 acres. Because of potential and actual damage to forests and forestry operations from these fires, government and private forestry organizations spend nearly $100,000,000 each year in preventing, preparing for, and suppressing them. In addition, many thousands of acres of woodland are intentionally burned every year under controlled conditions to accomplish some silvicultural or other land-use objective (Fig. 1).

Combustion of forest fuels. Successful control of forest fires depends on adequate knowledge of the combustion process and the environmental factors which influence it. In view of the importance of combustion to man, it is surprising how little is known about the fundamental physicochemical processes involved. This is especially true of unconfined fires.

Combustion of forest fuels is an exothermic oxidation reaction requiring for its initiation the application of sufficient heat to a suitable body of fuel in the presence of oxygen (*see* COMBUSTION). These three necessary and sufficient conditions can be considered to constitute three legs of a fire triangle. All three must be present for a fire to burn; removal of one will extinguish the fire.



Fig. 1. (a) Closed jack pine (*Pinus banksiana*) cones requiring heat to open. (b) Prescribed burning after seed tree cutting to open jack pine cones for natural regeneration. (Photographs by W. R. Beaufait)



Fig. 2 Fire control foreman checks the weather reading 3 times a day. (U S Forest Service)

Determinants of forest fire behavior. The manner in which fuel ignites, flame develops, and fire spreads and exhibits other phenomena constitutes the field of fire behavior. Factors determining forest fire behavior may be considered under four headings: attributes of the fuel, the atmosphere, topography, and ignition.

Fuel Live and dead vegetation in the natural forest complex constitutes the fuel for forest fires. The amount of moisture in relation to the heat content of the fuel determines its flammability. Moisture in living vegetation varies with species and with time of year Moisture content of dead fuels, the more important component of the fuel complex, responds to precipitation, relative humidity, and to a lesser extent, temperature It shows important daily fluctuations. *See* WOOD PHYSICS.

Rate of combustion and heat output are also determined by quantity of fuel, fuel arrangement, and the thermal and chemical properties of individual fuel particles *See* FUEL.

Atmosphere. Major direct effects of the atmosphere on fire behavior are through variations in oxygen supply and in flame angle caused by wind, and the stability of the atmosphere (*see* ATMOSPHERE; HUMIDITY) Secondary effects on combustion rate are through air and fuel temperature, and air relative humidity (Fig 2).

Topography. Heated air adjacent to a slope tends to flow up the slope, creating a chimney effect Unburned fuels above a fire advancing up a slope are closer to approaching flames because of the angle of the slope. Both of these effects act similarly to wind on flat ground in increasing rate of fire spread.

Ignition. An igniting agent of sufficient intensity and duration is necessary to start a forest fire. These agents may be man-caused, such as cigarettes or trash burners, or natural, such as lightning. During 1957, 68% of the wild fires in the United States were started accidentally by man, 23% were of incendiary origin, and 8% were started by lightning. These percentages vary widely

by region, however. In the 10 Rocky Mountain states, 71% were started by lightning and 2% were incendiary; in the 11 southern states, 2% were started by lightning and 38% were of incendiary origin.

Once a fire has started, its spread is by successive ignition of unburned fuel adjacent to the burning area. This fuel is brought to ignition temperature by heat radiated and convected from the flame front. Generally, fuels must be very close to or even enveloped by flame or superheated gases in the convection column before igniting. This mechanism results in more or less continuous advancement of a fire. Fire spread may also be caused by spotting, that is, ignition by burning brands carried outside the fire area by wind and turbulence.

Fire behavior. A forest fire may burn primarily in the crowns of trees and shrubs, a crown fire; primarily in the surface litter and loose debris of the forest floor and small vegetation, a surface fire; or in the organic material beneath the surface litter, a ground fire. The most common type is a surface fire (Fig. 3). Crown fires are usually accompanied and sustained by a surface fire, although occasionally a fire will burn over an area twice, once as a surface fire, then later as a crown fire.

Since fire is a self-sustaining chain reaction, its behavior is determined not only by existing environmental conditions but also by the conditions produced by the fire itself. For example, the vertical convection column produced by the fire creates indrafts. These indrafts increase the rate of combustion which leads to stronger vertical convection and consequently stronger indrafts. Because of this feedback, a slight change in environmental conditions, such as fuel becoming slightly drier, may produce a large change in fire behavior. The feedback principle holds for all levels of fire behavior. See CONVECTION (HEAT).

Occasionally fires exhibit erratic phenomena not readily explained on the basis of known weather-fire relationships. These phenomena include flash-overs, apparently in accumulations of unburned volatile matter, and extreme rates of spread under conditions thought to be comparatively safe. Spectacular increases in fire intensity in a short time are usually termed blow-ups.

Fire danger rating. In order to satisfy the needs of fire control organizations for estimates and predictions of potential protection work load, numerous systems have been developed which attempt to integrate selected determinants of fire behavior into numerical indexes of fire danger. These expressions of fire danger are related to such fire phenomena as rate of fire spread and probability of fire occurrence.

Most of these so-called fire-danger meters use fuel moisture content and wind speed as the fundamental variables. Among the methods used to estimate fuel moisture content are standardized sticks exposed in the forest environment and weighed periodically to determine changes in moisture content; relative humidity; state of annual vegetation, live or dead; and amount of and time since last precipitation.

Fire seasons. Because of annual cycles in precipitation and in growth of vegetation, most sections of the country have a well-defined fire season during which most forest and wild-land fires occur. This fire season varies widely in length and duration from two short seasons in the Northeast, spring and fall, to frequently a year-long season in southern California. Figure 4 shows the normal peak fire seasons for the United States.

Fire suppression. Because of the damage done by fire to forests and wild lands in terms of timber, watershed, range, and aesthetic values and because of the threat to human life, efforts have been made since the turn of the century to prevent and control wild fires. Although protection of wild lands from fire is a more recent development than protection of urban areas, the two fields share common fundamentals. Development and application of these fundamentals differ in the two areas; but these differences are largely in degree rather than in kind, and many of them stem from economic considerations rather than from the nature of fire.

Simply stated, the problem is to prevent potentially damaging fires from occurring and to suppress those that do occur efficiently with minimum



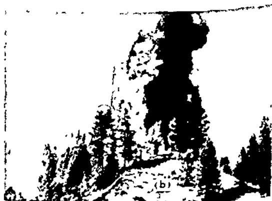


Fig. 3 (a) Surface fire. (b) Crown fire (U.S. Forest Service)

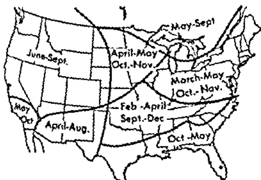


Fig. 4. Normal peak fire seasons in the United States.

cost. Certain elements are common to all solutions: fires may be prevented; those that start must be detected; this information must be communicated to the suppression forces; those forces must be transported to the fire; specific suppression methods must be employed in a strategic manner to effect control of the fire.

Prevention. Emphasis can be placed either on keeping igniting agents away from forest fuels (risk reduction) or reducing or eliminating the flammable materials (hazard reduction). Efforts directed toward reducing the incidence of man-caused fires encompass law enforcement and educational activities. The latter include campaigns aimed at the mass audience, for example, the "Smokey Bear" and "Keep Green" programs, and efforts directed at individuals, by forest patrolmen.

Research since 1950 shows promise in developing means for reducing lightning-caused fires by reducing cloud-to-ground lightning. The proposed technique is to seed developing lightning storms with silver iodide crystals causing the clouds to dissipate before reaching the lightning stage. *See* WEATHER MODIFICATION.

In areas of intensive use, many forestry activities are aimed at reducing fuel accumulations in order to reduce the chance of a fire starting or to reduce

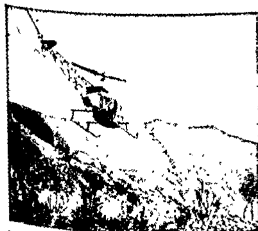


Fig. 5. Helicopter transporting fire fighters to scene of action (U.S. Forest Service)

the severity of one that might be started. Hazard reduction activities include road-side burning, slash disposal, controlled burning of forest areas to reduce fuel accumulations, and certain silvicultural practices. *See* SILVICULTURE.

Detection. Although many forest fires are detected and reported by local residents and by transient forest users, primary reliance for their detection is placed on specialized detection systems. Most fire-control agencies use lookout towers manned during the fire season by trained observers. These are often supplemented during critical periods by aerial patrol, using light planes or helicopters (Fig. 5).

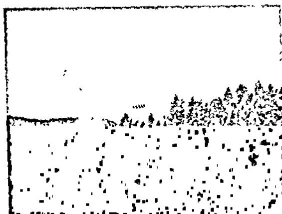
Electronic devices and methods developed during and after World War II have received some consideration and experimentation for detecting and locating fires and lightning storms that might start fires. These include television, infrared detectors, radar, and sferics (atmospheric characteristics detectable by radio methods). Although some of these have shown considerable promise, none has been adopted for routine operational use.

Communication. Rapid and accurate communication is essential not only to the detection network but also to the forces engaged in actual suppression work. Land-line telephones are being rapidly supplanted or supplemented by radios. In many organizations all motorized units are equipped with two-way radios. As the size of portable, self-contained equipment is reduced, more and more supervisory personnel on actual fires are also being equipped.

Transportation. Emphasis on fast attack of discovered fires has steadily increased over the years. Many fire-control organizations, both urban and forest, divide their protection area into zones of allowable attack times. These are based on times shown by experience as necessary for successful control to acceptable standards. Location of suppression forces and methods of transportation are planned to meet these objectives.

Because of the great distances involved, forest-fire-control organizations have placed special emphasis on increasing the speed and mobility of their forces. Most personnel and equipment arrive at the majority of fires by a combination of ground-transportation methods, including the most elementary of all, walking. Therefore, in areas of high value and high fire incidence an extensive network of roads and trails is usually maintained. The development of reliable, light-weight gasoline engines has made possible the use of trail scooters and motorized equipment carriers. These require trails intermediate in quality between horse and foot trails and the more expensive truck trails.

In remote areas with relatively low and widely scattered fire incidence, the airplane and helicopter are the primary means of delivering both men and materiel to the scene of a fire (Fig. 6). Even in areas with good road and trail networks, it is often cheaper and faster to supply suppression



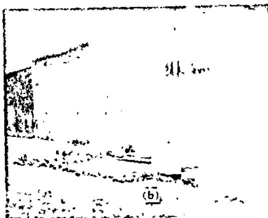


Fig 6 (a) Two fire fighters descending to a forest fire. (b) Air drop of fire fighting equipment showing drop location marker (X). (U.S. Forest Service)

forces by air. Successful parachute-jumping by fire-fighters, started by the U S Forest Service in 1940, demonstrated the feasibility of war-time use of paratroops Postwar development by the Forest Service of techniques and equipment for delivery of men and materiel by parachute has led to increased use of airborne fighters in the forests of the western United States.

Helicopters are being used more and more for transportation of men and equipment as their capacity and service ceiling are raised Because of their ability to land on hastily prepared heliports and because the men so delivered require no special training such as that required by parachute jumpers, helicopters are being used increasingly in all phases of fire control work.

Suppression methods. As indicated, the only methods by which a fire can be suppressed are by removing one or more legs of the fire triangle. Thus, a fire may be robbed of its fuel or its oxygen supply, or enough heat may be removed to stop the

may be done by digging a trench to mineral soil ahead of an approaching fire of sufficient width to prevent the fire from crossing (Fig. 7). Since this may require a line of considerable width (some evidence is available that this width is approximately equal to the square of the height of the flames), it is customary for the unburned fuel between the

terrain, soil character, and type of vegetation. Organized crews with hand tools, such as axes, shovels, rakes, and other specially constructed tools, are used in many areas, especially those too rugged or remote to permit use of mechanized equipment. Specially developed motor-driven flails, similar to garden-tractor rotary tillers, are sometimes employed, although their usefulness is limited to terrain and soil types on which they can be maneuvered satisfactorily. Bulldozers are used extensively, and tractor-drawn plows are standard line-construction equipment in the relatively flat terrain of the southeastern United States.

Since 1954, some use has been made of chemicals to prepare a nonflammable fire line. A water-base slurry is sprayed in a line on surface fuels from which a fire may be started to produce a wide fuelless line in the same manner as are dug lines. More commonly, the slurry is dropped in bulk from a low-flying airplane in a line just ahead of the advancing fire. The fire, on reaching the coated vegetation, either stops completely or is reduced in intensity sufficiently to enable ground crews to work safely and effectively on the fire front.

Occasionally existing roads or previously prepared firebreaks are used as control lines, usually as a line from which a counter-fire is set. Natural firebreaks, such as rocky ridges or rivers, are used similarly.

The second major method used in suppressing fires is to remove enough heat to break the chain reaction. Water, with its high specific heat and heat of vaporization, is an ideal substance for this purpose. Long the favorite with urban fire control organizations, water is becoming more popular with forestry organizations as techniques and equipment for handling it are improved. An extraordinarily large amount of burning material can be extinguished with a small amount of water when it is used with maximum efficiency. For specialized operations, wetting agents are added to water to increase its effectiveness.

Hand-operated pumps with a 5-gal tank carried on the back are most satisfactory when used on fires in light fuels, for suppressing small spot fires, and for mop-up, the tedious work of extinguishing burning fuel after the fire has been contained within a fire-proof boundary.

Gasoline-operated pumps are used extensively wherever water supplies are convenient or road networks permit the use of tank trucks. Portable pumps are available that can be carried by one



Fig 7. (a) Digging a trench ahead of an approaching fire (b) Constructing a fire line to stop the advancing fire. (c) Bulldozer making fire lane near edge of fire. (d) Using a tank truck to suppress fire. (e) Fighting fire with hose attached to portable pump. (U.S. Forest Service)

man, making hose lays of several thousand feet or even several miles feasible. New techniques have been developed for laying hose rapidly from helicopters, cutting time required to lay a mile of hose from hours to minutes, even seconds. For use in inaccessible regions, small units including a pump, a 50- or 100-gal tank, and several hundred feet of hose have been developed that can be delivered by helicopter to the scene of a fire.

The airplane has also been used since about 1952 to make direct water drops on fires. Cargo airplanes modified to carry up to several hundred gallons of plain water or a fire-retardant chemical mixture are used to cascade the water or chemical on a portion of the fire, usually the leading edge. Initial success with this technique has led to its increasingly widespread use. Less widely used but finding specialized application is another method that utilizes waterproof paper bags holding about 3 gal of water which are dropped on small fires by using dive-bombing tactics.

Little direct use is made of the third possibility of removing the oxygen supply to a fire, although water has this effect to some extent and soil is sometimes used to smother flames.

Suppression strategy. The manner in which the various suppression techniques are utilized on a fire is determined by the fire boss, who is comparable to the fire chief in urban organizations. It is his responsibility to determine the strategy appropriate to the particular fire. Depending on the size of the fire, he may have a few men or a few hundred, and occasionally more than a thousand, to assist him in carrying out his strategy. Forestry organizations maintain intricate paper organizations and large caches of fire-fighting materiel which can be mobilized rapidly in time of urgent need. See FOREST AND FORESTRY; FOREST CONSERVATION. [W.E R.]

Bibliography: K. P. Davis, *Forest Fire Control and Use*, in press; A. D. Folweiler and A. A. Brown, *Fire in the Forests of the United States*, 1953.

Forest mapping

That aspect of forestry dealing with the preparation of specialized maps showing the distribution and conformation of the individual forest stands,

specialized techniques and training for the characterization and delineation of forest stands (see CARTOGRAPHY; SURVEYING). Since foresters are usually charged with the responsibility of managing large areas of relatively low-value forested land, the preparation of forest maps is a basic requirement in forest management See SILVICULTURE.

Ground survey. Before the advent of low-cost aerial photography, forest maps were prepared on the ground by survey parties. Ground methods are still widely used for mapping small areas and areas of high value. If available, existing maps of the desired scale are used as a base upon which the forest information is compiled. If not, base maps must be developed by standard mapping procedures (see MAP DESIGN). The forester then traverses the area, usually along a series of parallel courses. These lines are ordinarily run in a cardinal direction, north-south or east-west, in level country, and across the topography at right angles to the major drainages in hilly country (see TOPOGRAPHIC SURVEYING AND MAPPING) Direction is determined from a surveyor's compass mounted on a staff or similar simple compass, and distance is measured by a steel tape or by pacing when a map must be prepared quickly by experienced foresters. Along each course, the mapper notes the point at which each succeeding forest stand is entered and left. He describes each stand with a series of code symbols and sketches the stand boundary for a short distance on either side of his course. If the courses run are intervisible, the map will be completed when the entire area has been traversed. Most commonly, however, the lines are 400–1500 m apart, and some additional surveying is necessary to connect the segments of forest stand boundaries mapped along the lines. Transits and similar instruments are seldom used in forest mapping except in the location of land ownership boundaries and elsewhere where precise location is important. See TRANSIT (ENGINEERING).

Aerial survey. Beginning after World War I and attaining widespread acceptance after World War II, aerial mapping techniques have greatly facilitated low-cost, high-accuracy forest survey. Direct sketching from low-flying, slow-speed aircraft has been used extensively in the past, especially in Canada, and is still employed occasionally for special projects. However, helicopters are best for this work.

Most forest maps are developed from aerial photographs (see AERIAL PHOTOGRAPH). For forestry purposes, aerial photographs are generally taken with special mapping cameras pointed vertically at the ground from elevations of 2000-6000 m. Standard negative size is 23 × 23 cm. Each successive exposure is taken so as to rephotograph 60% of the area covered by the last previous exposure. This overlap permits three dimensional stereoscopic study of any two adjacent pictures in the same flight line, and the determination of both horizontal



Scanning stereoscope for photointerpretation in forest mapping. (N. V. Optische Industrie)

position and vertical elevation of points pictured on the ground.

Most forest photography is taken with black and white film sensitive to the visible spectrum (panchromatic), but infrared-sensitive film is valuable for separating conifers from broad-leaved species, and color film is useful for special purposes, such as identifying diseased and damaged trees or for distinguishing between species. Photographs having scales of 1:10,000 to 1:20,000 are most commonly used These scales are large enough to present sharp images of individual trees under stereoscopic magnification, and yet small enough to cover a substantial area in a single photograph See PHOTOGRAPHY.

In the office, the map position of selected points visible on three or more overlapping photographs is determined by the resection of lines extended from the centers of each photograph through the image of that point on that photograph. By this radial-line triangulation procedure, or by similar techniques based upon mathematical analysis or utilizing special instruments, the map position of the center of each photograph and the selected points can be plotted. Map detail is then transferred from the photographs to the map by special photogrammetric devices (see PHOTOGRAMMETRY). For rapid work, the image of the photograph can be projected or otherwise superimposed on the map and the detail traced. To eliminate displacement errors, however, it is necessary to view the photographs in pairs stereoscopically and to correct for the effects of elevation of points on the ground and the usual slight amount of tilt in the photographs themselves.

Forest types. On almost all forest maps, the forest types are delineated. These are areas of forest more or less homogeneous with regard to tree species, and are usually named after the most numerous or most important components, for example, birch-beech-maple type, Douglas fir–western hemlock type, or mahogany type. The type may be composed of one species, as aspen in the lake states, or of hundreds, as in the tropical rainforest. To be delineated on the map, the type must occupy a certain minimum area, for example, 1 hectare on a map of 1:10,000 scale. The most important forest types in North America exclusive of Mexico have been given standard names by the Society of American Foresters.

Stand description. In addition to the forest type, the stand is usually also classified according to such characteristics as its density, tree size, stand volume, and age. In the United States, emphasis is placed upon tree diameter. Thus, saplings are classified as having stems less than 5 in. (12.7 cm) in diameter at breast height (4.5 ft or 1.37 m from the ground); pole timber from 5 to 9 in. (12.7–22 9 cm); and sawtimber above 9 in. In Europe, the description preferred is in terms of stand age class, by 10- or 20-year groupings, and density to the nearest 10%, with 100% representing maximum stocking conditions. With modern aerial photographic techniques, however, there is a trend toward stand description in terms of relative crown closure and tree height because these can be evaluated through stereoscopic photointerpretation.

Site quality. For forest management purposes, the ability of the land to produce forest is as important as the present forest on the land In current forest mapping, much emphasis is placed upon the mapping of this site quality, which is the sum total of the soil and climatic factors affecting tree establishment and growth. Since it is impractical to assess all the site factors, site must be indexed according to the observed growth of trees, or to one or more key soil characters. Site index, the height of free-grown trees at 50 or 100 years, is widely used where the desired trees are already growing on the site. In northern Europe and elsewhere, the presence or absence of certain indicator plants on the forest floor, such as *Cladonia, Calluna, Vaccinium,* and *Myrtillus,* has proved an excellent guide to forest site quality. Among the soil characters best related to site quality are parent geological material, the chemical fertility of the soil, the size of the soil particles, the depth of the portion of the soil usable by tree roots, the depth to the water table, and the water-holding capacity of the soil (*see* FOREST SOIL). [S H.S.]

Bibliography: *See* FOREST AND FORESTRY.

Forest mensuration

The branch of forestry concerned with the measurement of trees and products derived from them. Measurements may be made at any time during the growth of the tree or at the completion of any of the processes involved in the harvesting and manufacture of forest products (*see* LUMBER MANUFACTURE). Tools and techniques used are determined largely by the stage of growth or manufacture. This article will begin with the seedling and end with the harvested product.

Seedlings. The most common measurements, or descriptive statistics, for very small trees are based on per cent survival for planted seedlings or on an index of stocking for natural reproduction. Per cent survival is obtained from a direct count of living seedlings using the known number planted as a base. Various indexes of stocking are employed for natural reproduction. A simple count may be made and expressed as a number of seedlings per acre, but this statistic does not account for distribution. Sizable areas devoid of seedlings may be present even when counts are high. A more useful index, commonly called per cent stocking, is obtained by dividing the area into mil-acre (1/1000 acre) squares and counting the number containing one or more trees.

Saplings. Measurements on trees taller than 4.5 ft but too small to be merchantable are usually confined to diameter, height, and density or stocking. Diameter is almost invariably measured at breast height (4.5 ft above the ground level) in inches and, for refined work, tenths of inches. Three instruments are used for measuring: calipers, diameter tape, and the Biltmore stick.

Tree calipers consist of two arms perpendicular to a graduated bar (*see* CALIPERS). One arm is fixed at the zero point; the other slides along the bar parallel to the fixed arm (Fig. 1). In use, the arms are placed in contact with opposite sides of the tree and the diameter is read from the scale on the bar. Correctly adjusted calipers offer a fast and accurate means of measuring the diameters of small trees. However, the bulk of the instrument

ble steel tape that may be wrapped about a tree (Fig. 2). Each major division is 3.1416 or π inches apart; hence the tape reads directly in inches of diameter. Although slower, it is at least as accurate as calipers, and has the added advantage of com-

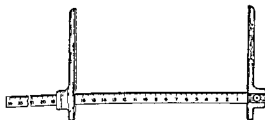


Fig. 1. The tree caliper. (*From D. Bruce and F. X. Schumacher, Forest Mensuration, McGraw-Hill, 1935*)

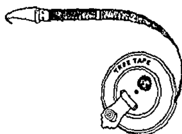


Fig 2 Tree diameter tape. (Forestry Suppliers Inc., Jackson, Mississippi)

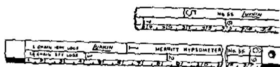


Fig 3. Biltmore cruiser stick. (Forestry Suppliers Inc., Jackson, Mississippi)

pactness with no adjustments to get out of order.

The Biltmore stick is a bar graduated so that diameters may be read by sighting over the scale when the stick is held at arm's length tangent to the tree and perpendicular to a line drawn from the eye of the user to the pith of the tree (Fig. 3). Fastest but least accurate of the three, it is commonly used by timber cruisers (estimators) but never where precise measurements are required. Slight variations in positioning the stick against the tree cause large errors in diameters read from the scale.

The heights of small trees are usually measured with a graduated pole or rod, but a form of indirect measurement is necessary for trees taller than 15 or 20 ft, as the length of rod required becomes unwieldy. A number of instruments have been devised for this purpose, all based on the principle of triangulation. See SURVEYING.

The most widely used hypsometer is the Abney level, a small hand level with a bubble and scale for measuring vertical angles (Fig. 4). Ordinarily, the scale reads, not in degrees, but either in per cent slope or, in the case of the topographic arc, in number of feet of rise in 1 chain (66 ft). To measure heights with the instrument, two sights are taken, one at the base of the tree, the other at the tip of the crown. The difference in the readings on the per cent scale gives height in feet directly if the user is located 100 ft from the tree. If the topographic scale is being used, direct readings of height are obtained at a distance of 66 ft from the tree. Either scale may be employed at other distances but some simple arithmetic must then be performed to transform the scale readings to height.

Instruments similar in principle to the Abney level but employing pendulums instead of a bubble vial are the Forest Service hypsometer and the Haga altimeter (Fig. 5). The last named has a number of scales mounted on a revolving bar. The scales are calibrated to read height in feet at different distances from the tree.

A number of simple instruments based on the principle of similar triangles are also used to measure height. The most common is the Merritt hypsometer, which is simply a stick graduated to read height directly at a fixed distance from the tree. The stick is held at arm's length perpendicular to the ground with the bottom end aligned with the base of the tree. The height reading is taken where a line of sight from the eye to the tip of the tree crosses the scale on the stick. The graduations for the Merritt hypsometer are often placed along one edge of a Biltmore stick, forming an instrument that may be used for both height and diameter estimates.

Where an index of stocking or density is required for saplings, a simple count of the number of trees per acre may be employed or diameter measurements may be transformed to areas of equivalent circles and totaled to obtain basal area, a measure of density defined as the cross-sectional area of the tree stems 4.5 ft above the ground, usually expressed in square feet per acre.

Basal area may also be obtained directly, without diameter measurements, by means of an angle gage, which may be an optical device such as a prism, or merely a stick with a peephole at one end and an upright bar at the other. To measure basal area with the gage, the observer stands on a fixed point and counts all the trees around him that subtend an angle larger than that for which his gage is calibrated. The count, when multiplied by a converting factor appropriate for the critical angle of the gage, gives basal area in square feet per acre. Usually, counts are taken at a number of points within the stand and averaged to obtain the final estimate.

Cordwood-sized trees. Once trees have reached cordwood size, most species are merchantable and tree volume becomes of interest. The volume of such trees is ordinarily measured either in cubic feet or cords. The latter is a unit of stacked volume and is defined as equivalent to a rick of wood 4 ft wide, 4 ft high, and 8 ft long.

In order to avoid complications caused by varying standards of merchantability for different species at different times and places, the total cubic volume of the stem may be measured. The tip or

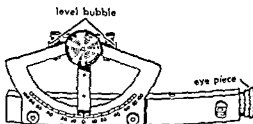
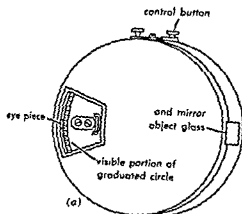


Fig. 4. The Abney level. (From D. Bruce and F. X Schumacher, Forest Mensuration, McGraw-Hill, 1935)



(b)

Fig 5. (a) The Forest Service hypsometer (from D Bruce and F. X. Schumacher, *Forest Mensuration*, McGraw-Hill, 1935). (b) Haga altimeter (Forestry Suppliers Inc., Jackson, Mississippi).

upper stem of the tree is included, but the roots and stump are almost invariably ignored. The bark of the main stem may or may not be included The adjective *rough* appended to the volume statistics indicates that bark is included; similar use of the word *peeled* indicates that the converse is true.

Volumes measured in cords include only the portion of the tree stem that is merchantable at a given time and place. Often the volume of the tree is first measured in cubic feet and then converted to cords by application of the appropriate multiplier. Volumes given in cords may or may not include bark; again the adjectives rough and peeled designate which is the case.

The volumes of standing trees are usually obtained from tables prepared in advance and based on felled trees. Volume tables differ chiefly in the independent variables employed; that is, they may be constructed to give volume for a known diameter at breast height, or for a known diameter and

Sample volume table for red spruce in the Northeast*

Diameter at breast height, in	Volume, ft³			
	40-ft tree	50-ft tree	60-ft tree	70-ft tree
10	10 00	12 40	14 70	17 60
11	12 00	14 80	17 70	21 00
12	13 70	16 90	20 20	23 70
13	15 60	19 40	23 60	28 00
14	17 50	22 00	26 50	31 70
15		24 00	29 00	34 00
16		25.00	30.10	36 20

* Volumes are for entire stem above a 1-ft stump, from H H Chapman and W H Meyer, *Forest Mensuration*, McGraw-Hill, 1949

height, or for a known diameter, height, and form

top of the first log (16.3 ft above stump height) to the diameter outside the bark at breast height.

Tables based on diameter alone are called local

local and standard tables are commonly made for individual species, but tables entered with a measure of tree form can be applied to a number of species.

A number of methods of making volume tables are in use, but all begin with the collection of measurements on sample trees. The volumes of these sample trees are computed, either graphically or by means of formulas, from diameter measurements taken at intervals along the stem.

It is chiefly in the process of averaging that the various methods of table construction differ. Formerly, free-hand curves and alignment charts were widely used, but in recent years mathematical fitting of curves by the process of least squares has become the usual practice. Examples of equations used for this purpose are the logarithmic equation

$$\log V = a + b \log D + c \log H$$

the combined variable equation

$$V = a + bD^2H$$

and the volume-diameter ratio form

$$V = a + bD + CDH + dD^2 + eH + fD^2H$$

where V is volume, D is diameter, and H is tree height in all three equations.

In using a volume table to obtain the volume of timber in a particular stand, the diameters and

ume table. Later computations are then considerably simplified; the volume of each individual tree need not be obtained. Instead the number of in each class is multiplied by the table vol

that class and the results summed to obtain total volume for the stand.

On large areas, the usual practice is to measure only a sample of the trees for volume. The sample may be chosen (1) by using the individual trees as the sampling unit, (2) by using plots or strips as the sampling unit, or (3) by using sample points in connection with an angle gage.

In the first method all trees on the area are counted, and a portion, chosen systematically or at random, is measured as required by the volume table. The total volume of the sample trees is then computed from the table and the results multiplied by the inverse of the sampling fraction to obtain total volume for all trees on the tract.

When plots or strips are used, a fixed portion of the ground area rather than a fixed fraction of the number of trees is sampled. Plots or strips are distributed over the area, again either systematically or at random, and all trees on the sampling units are measured for volume. Sample volume and total volume are then obtained as described for the preceding case.

If cruising is done with an angle gage, the field procedure is quite similar to that described for plots and strips. However, the tally obtained gives, after multiplying by the angle gage factor, not number of trees but basal area by the tree classes recognized in the timber cruise. The total basal area for each class is converted to number of trees by dividing by the average basal area for the class and volume computations and then proceeding as described for the two preceding cases.

Systematic location of plots, sampling points, and strips is the practice most commonly used by foresters and timber cruisers. However, in recent years the use of random samples has increased. Usually stratified random samples are employed, using aerial photographs as the means of delineating the strata (see AERIAL PHOTOGRAPH; FOREST MAPPING). More complicated survey designs embodying ratio estimates, subsampling, cluster sampling, and sequential sampling are just beginning to come into prominence in forestry.

Sawtimber-sized trees. Measurements of height, diameter, basal area, and volume are commonly gathered for sawtimber-size as well as cordwood-size trees. Methods of preparing volume tables and inventory procedures are much the same as those described for cordwood-sized trees. However, volumes are usually measured in board feet according to some standard log rule rather than in cubic feet or cords.

The board foot (fbm) is the basic unit of measure for lumber and is defined as a piece of wood 1 in. thick, 12 in. wide, and 12 in. long, or its equivalent. The dimensions apply only to rough lumber, however, and may be reduced by surfacing and, in the case of conifers, by shrinkage on drying. A board foot is not $^{1}/_{12}$ ft³; part of the cubic-foot volume of a log is lost in sawdust, edgings, and slabs when lumber is sawed.

Log rules estimate the amount of lumber that can be cut from logs of varying diameters and lengths. Rules in common use include the International, Scribner, and Doyle. The volume of a log 16 ft long and 16 in. in diameter inside the bark at the small end is International 180, Scribner 160, and Doyle 144 fbm. The International rule closely approximates the lumber tally obtained by a careful sawyer and is widely used for research purposes. The Scribner rule, rounded to the nearest 10 fbm and called the Scribner Decimal C, is employed by the United States Forest Service in sales from the national forests. The Doyle rule, the oldest of the three, is widely used in commercial practice, but this rule seriously underscales small logs and overscales large ones.

The Scribner rule is a diagram rule and was first tabulated by drawing on circles of varying size rectangles representing boards that might be sawed from logs having the same diameters as the circles. Both the Doyle and International rules are formula rules. The equation for the Doyle rule is

$$V = \frac{(D-4)^2 L}{16}$$

where V is volume in board feet; D is the diameter of the log at the small end in inches, and L is the length of the log in feet. The formula for the International rule is

$$V = 0.199D^2 - 0.642D$$

for a 4-ft log. Logs longer than 4 ft are treated in 4-ft sections with an increase in diameter of ½ in for each additional 4 ft in length.

Products. Many forest products such as posts, mine props, ties, poles, and piling are sold by the piece. Therefore, measurement is primarily a process of counting and classification. Trade associations promulgate specifications for the different classes of ties, poles, and piling. The specifications define the species acceptable, minimum and maximum size, defects allowable, and the amount of crook and sweep (curvature) permitted.

Pulpwood, fuel wood, and small specialty bolts (short logs) are commonly measured by the cord (see WOOD FIBER PRODUCTS). Deductions may be made for voids, spaces left in loosely stacked wood, and for knotty or crooked wood. Recently some mills have begun to buy pulpwood by weight. Where this practice is followed, measurement is simple, quick, and less subject to controversy.

Sawlogs, veneer logs, and large bolts are usually bought and sold by the board foot according to some standard log rule. Both the log rule employed and scaling practice vary widely. Measurement is almost invariably made at the small end of the log, but that is the extent of uniformity. National forest scalers take the average diameter of logs whose cross section is not circular, and round off the measurements to the nearest inch. Common commercial practice, on the other hand, is to measure

the smallest diameter and to round off the measurements to the nearest lower inch. Further variation is found in the method of scaling long or tree-length logs and in allowances made for sweep, crook, rot, and other defects.

Growth and yield. In addition to the measurement of existing trees and products, the measurement of past growth and the prediction of future growth form an important part of forest mensuration. Past growth is best determined by repeated inventories of trees and stands, and this practice has recently become more common in the United States. Often the mass of information collected from such inventories is summarized and analyzed on punched-card machines.

In the absence of repeated measurements, past diameter growth of individual trees may be determined by means of an increment borer, an instrument that removes a core of wood from the tree (Fig 6). The number of annual rings on the core may be counted and the widths measured to obtain radial growth inside bark. *See* WOOD (ANATOMY AND IDENTIFICATION); XYLEM. The radial growth must then be doubled and an allowance made for bark growth in order to get diameter growth exclusive of the bark (*see* BARK).

Past height growth of species, such as white pine and red pine, that form distinct annual nodes bearing branches may be obtained from the length of the internodes. The height growth of species that do not form such nodes may be obtained by stem analysis, a process that requires sectioning the tree and counting the annual rings in the end of each section. However, this process is expensive and is seldom employed except for research purposes.

The future growth of all-aged stands of trees is usually predicted by a process known as stand projection. A stand table giving the existing number of trees is prepared and local volume tables are applied to obtain present volume. The stand table is then projected into the future using past growth determined from increment cores as a guide and the volume table again applied to obtain a future volume. The volume so obtained must be reduced for expected mortality, and the uncertainty of the proper allowance to be made is the major weakness of the procedure.

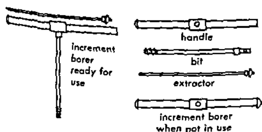


Fig. 6. Swedish increment borer. (From H. H. Chapman and D. B. Demeritt, *Elements of Forest Mensuration*, J. B. Lyon, 1932)

Growth predictions for even-aged stands are generally made on the basis of how similar stands have grown in the past. For a number of species, yield tables are available that list volumes that may be expected for normal (fully stocked) stands at various ages on various sites. Often, however, stands for which growth estimates are required are more or less heavily stocked than the yield-table standard. Several arbitrary adjustments for nonnormal stocking have been developed and used; however, more satisfactory predictions are obtained from tables that either give information on probable changes in density of stocking or include stocking as one of the independent variables. *See* FOREST AND FORESTRY. [C.M.F.]

Bibliography: D. Bruce and F. X. Schumacher, *Forest Mensuration*, 3d ed., 1950; H. H. Chapman and W. H. Meyer, *Forest Mensuration*, 1949.

Forest seeding and planting

The establishment of a new forest by seeding or, more generally, by the planting of nursery-grown trees from 1 to 4 years old, with tops usually from 4 to 12 in. tall and roots about 8 in. long

Reforestation had its origin in Germany when in 1368 several hundred acres of devastated land were direct seeded to a number of species of conifers. By the mid-eighteenth century, artificial reforestation was well established as a forest practice in Germany and had spread to Austria, Switzerland, and France.

Much credit for stimulating interest and providing successful examples of tree planting is due several Germans who came to the United States in the latter half of the nineteenth century. Among these was Carl Schurz, who in 1877 became the first Secretary of the Interior of the United States and who urged tree planting as part of a program of scientific forest management. Other leaders were B E. Fernow, who in 1886 became the first chief of the Federal Division of Forestry, and Carl Alwin Schenck who came to America in 1895 to serve as forester on the Biltmore estate near Asheville, North Carolina. As early as 1897 Schenck supervised the planting of 500,000 2-year-old white pine seedlings in the Biltmore forest. In this same period a few smaller-scale plantings of pine were made in Michigan and other states (Fig. 1). Extensive tree planting was not done in the United States until the passage of the Timber Culture Act in 1873. Early plantings were made largely in prairie and plains areas

In 1902, two state nurseries were established to grow trees for reforestation, one at Saranac Inn, New York, and another at Mont Alto, Pennsylvania. That same year, under President Theodore Roosevelt, an ardent conservationist, a Federal nursery was started at Halsey, Nebraska, for sandhill afforestation.

The Civilian Conservation Corps, established by President Franklin D. Roosevelt, planted about 2,250,000,000 trees from 1933 to 1942. In this same



Fig 1. A plantation of white pine established in 1896, one of the oldest in Michigan, photographed in 1952

period, new tree-planting machines accelerated reforestation, especially in the South, on land owned by forest industries.

By 1957, 1,000,000,000 trees had been planted in New York State under the state-sponsored program. Coincidentally, on a nation-wide basis, 1,000,000,000 trees per annum were being planted by 1957, equivalent to about 800,000 acres. Georgia, the leading state, planted about 100,000 acres that year.

Purpose of tree planting. Trees are planted to produce lumber and structural timbers, and to provide pulpwood (Fig. 2) for such products as paper, roofing felt, rayon, plastics, and particle boards of various types (*see* CELLULOSE; FIBER, MAN-MADE; LUMBER MANUFACTURE; PAPER AND PAPER PRODUCTS; PLASTICS FABRICATION; WOOD FIBER PRODUCTS). Plantations may also furnish wood for posts, utility poles, paneling, veneer, furniture, boxes, pallets, and handles, as well as Christmas trees, turpentine, and pine oil (*see* WOOD CHEMICALS). Tree planting may produce numerous other benefits such as decreasing peak flow in floods, decreasing water and wind erosion, and furnishing food and cover for wildlife. *See* FOREST CONSERVATION; FOREST ECOLOGY; SOIL CONSERVATION; WATER CONSERVATION.

Seed collection, storage, and pretreatment. Most tree seed ripens at the end of the growing season; however, there are some exceptions, such as the cottonwoods, willows, and most elms which ripen their seed in spring. *See* SEED (BOTANY). Seed of others, such as rock elm, ripen in midsummer, whereas a few conifers, such as white spruce, may ripen seed in late August. *See* CONIFERALES.

Tree seeds may occur on the tree singly, or grouped in berries, pods, cones, or clusters of samaras. *See* FRUIT (BOTANY). They may be picked by hand from standing trees or recently felled trees, or, after ripening, from the ground. Squirrel hoards (caches) sometimes yield several bushels of seed-bearing cones of pines or other conifers. Seed of such trees as elms and junipers may be shaken from the tree and collected on canvas, plastic, burlap, or a cloth spread on the ground.

Nearly all seed of the important conifers used in reforestation, such as pines, spruces, true firs, Douglas firs, cypresses, cedars, junipers, and larches, may be stored for periods of several years (up to a decade or more with a minimum loss of viability) at 33–41°F in sealed containers, and at a seed moisture content of under 10% on a dry-weight basis.

Some tree seeds have no dormancy, or after-ripening requirement, and germinate promptly within a week or two after sowing. Scotch pine and jack pine are examples. Other tree species, especially certain legumes, have hard, almost impermeable seed coats which prevent water and oxygen from reaching the embryo for a period of many weeks or months (*see* LEGUME). To obtain prompt uniform germination, the seed coats of such species as black locust and honey locust are scarified (scratched) in sandpaper-lined drums, disk scarifiers, or other devices. Alternatively, the seed coats may be made more permeable by soaking them for a period of about 30 min in concentrated commercial-grade sulfuric acid, removed, drained, and thoroughly washed with clean water before sowing.

Some species of tree seed, such as white pine, have embryo dormancy and must undergo an after-ripening process at cool temperatures and in a moist condition to expedite certain chemical changes necessary before germination can take place. The seeds are stratified, or mixed with sand or peat, moistened, and placed in cold storage at 33–41°F (up to 50°F for some species) for several months before sowing in the nursery.

A few species have a combined hard seed-coat and embryo dormancy and require both scarification (or acid treatment) and cold moist stratification before sowing in the nursery.

Trees grown in nurseries usually require 1–4 years before they are large enough for reforesta-



Fig. 2. A 20-year-old red pine plantation in the Chippewa National Forest, Minnesota, with pulpwood (foreground) taken from thinnings.

tion planting *See* FOREST TREE NURSERY MANAGEMENT; PLANT GROWTH.

Planting site preparation. Planting sites, such as recently burned-over forest or brush land, may not require plowing or other ground preparation (Fig. 3). On rough, hilly, or mountainous land with considerable grass, herbaceous, or brush cover, a spot about 18 in. square is scalped off with a heavy hoe or mattock just before planting each tree. Where machinery can be operated efficiently, heavy brush plows, disks, or bulldozers are sometimes used to tear up the brush and make prepared lanes of exposed mineral soil in which to plant the trees On soddy land an ordinary one-bottom plow or a forestry plow may be used to open parallel furrows, from 6 to 8 ft apart, in which the trees are planted at about 6-ft intervals. On rolling land such furrows are placed on the contour to conserve moisture and reduce soil washing. *See* CONTOUR, TERRACING (AGRICULTURAL).

Chemical sprays of 2,4-D, of 2,4,5-T, or combinations of the two, applied by ground equipment or by aircraft, are used as a means of eliminating unwanted scrub tree or brush growth. Ammonium sulfamate applied by ground equipment is also useful for this purpose. *See* HERBICIDE.

Season of planting. Most tree planting is done at the beginning of the growing season while the trees are still dormant and can be moved more successfully. Planting in the southern United States, in southern Europe, and in Mediterranean lands is done in early winter.

Tree species to plant. The most common tree species planted in the northern and northeastern United States are red, eastern white, and jack pines and white, black, and Norway spruces; in the South, shortleaf, loblolly, slash, and longleaf pines are most important; whereas in the West, Douglas fir, ponderosa, western white, and sugar pines are important as well as Engelmann spruce. For descriptions of these trees see separate articles listed by common names. In European and Asiatic countries, planting is usually done with native species of pine, spruce, fir, cypress, cedar, larch, oak, ash, and other hardwoods. In Australia and South Africa outstanding results have been obtained with Monterey pine, introduced from California.



Fig 3 Burned-over pine land with scrub growth of aspen is a site suitable for reforestation.

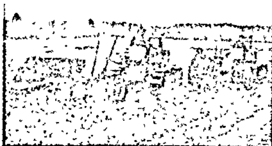


Fig 4 Two tractor-drawn tree-planting machines used in reforestation with pine trees in Michigan. The site is sandy, impoverished former farm land.

Machine planting. Fairly level to moderately undulating (rolling) land may be planted by machine. The planting machines are either 1-row or 2-row units and are generally pulled by a crawler tractor On fairly smooth terrain, wheel tractors can be used Two men and a machine can plant 10,000 or more trees per day (Fig. 4).

Hand planting. On steep or rocky land a spot of bare mineral soil is exposed and the trees are planted with a mattock, or more rarely with a planting bar at approximately a 6 × 6 ft spacing.

Planting in pots or asphalt bands. In subhumid and semiarid areas where frequent failures occur with bare-rooted stock, trees are sometimes grown in pots of clay, cement, or asphalt paper, and transported to the planting site in the container. The plant with its 2- to 4-lb cylinder or block of soil intact about the roots is then removed from the first two types of container and planted on the site. Asphalt bands (containers) are not removed because these permit entrance of water to roots, and they also deteriorate naturally

Seeding. For oak, walnut, beech, pine, fir, spruce, and a few other genera of trees, direct seeding is done occasionally by dropping seeds, or by drilling in furrows, prepared terraces, scalped spots, or recently burned-over areas. Aircraft and helicopters have been used for seeding on large burned-over areas. However, because of the frequent adverse results owing to drought, competing vegetation, and seed-eating rodents and birds, the use of direct seeding usually has not been successful as a large-scale reforestation measure.

Plantation care. Plantations commonly require removal of overtopping weeds, brush, or inferior trees with tools or chemicals. They may also require the use of sprays to protect against insects. Protection from disease is sometimes accomplished by removal of the alternate host plants Measures are often needed to prevent forest fires, and rodents as well as larger mammals may be controlled by use of rodenticides, repellents, or fences. *See* FOREST AND FORESTRY; PLANT DISEASE CONTROL; RODENTICIDE; TREE. [J.H.ST.]

Bibliography: J. W. Toumey and C. F. Korstian, *Seeding and Planting in the Practice of F* 1942.

Forest soil

The natural medium for the growth of roots of trees and associated forest vegetation. This relationship with forest vegetation gives rise to characteristics which distinguish forest soils from other soils. The most obvious feature is the humus layer, which is peculiar to the microenvironment imposed by the forest (Fig. 1). The humus horizon determines the course of forest reproduction, influences soil moisture distribution, serves as a reservoir of nutrient elements … for … influ- … Organic remains of plants and animals of the forest are deposited near the soil surface and gradually decomposed. The course and duration of decomposition which determines …

Organic matter may accumulate on the ground surface with little or no mixing with the mineral soil. This form is called mor, raw humus, or ectohumus. In mull or endohumus, much of the organic matter is incorporated into the mineral soil and only coarse debris, such as twigs and petioles, remains on the surface (Fig. 2). Mor humus is usually associated with a cool, moist microclimate and predominantly coniferous vegetation growing on acidic parent material. In these situations there is a general lack of many of the animal organisms responsible for incorporating organic matter into the mineral soils, and the micropopulation tends to be dominated by fungi.



Fig. 1. Soil under an uncut forest is friable and porous, deeply penetrated by roots and infiltrated with organic matter. (USDA)

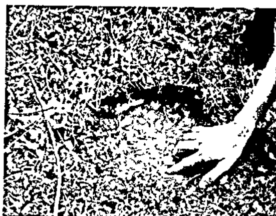


Fig. 2. Forest cover returns large quantities of organic matter to the soil each year. Depression in center shows depth of litter. (USDA)

The leaf litter associated with mull soils is usually richer in calcium and other nutrient elements and hence supports a larger variety of small animals and microorganisms than are found in a mor layer. A favorable microenvironment stimulates the activity of these organisms so that litter breakdown proceeds more rapidly. Earthworms, millipedes, and similar animals also partially digest the forest litter and incorporate it into the mineral soil. Simultaneously, a burst of bacterial and fungal activity rapidly converts the organic matter into somewhat more stable forms such as lignin and microbial remains. These processes result in a darkening of the upper part of the mineral soil.

Humus relations. The factors which govern humus formation are extremely complex, and their relative importance is not well understood. Mull formation appears to be associated with an abundance of biological activity, an adequate moisture supply, relatively high temperatures, and tree species having nutrient-rich litter. When organic matter accumulates instead of being incorporated into the soil, an intermediate form showing characteristics of both mull and mor results.

A definite relationship exists between humus morphology and the horizons which occur deeper in the soil profile. Many of the reactions responsible for profile development originate in the humus layer. Mor humus is often associated with pronounced leaching and horizons of accumulation, whereas in mull soils the profile transitions may be less distinct. Although the humus layer is sensitive to changes in the aerial environment, the deeper horizons usually are more stable. When the microclimate of a forest stand is modified by wind damage or silvicultural operations, the humus layer adjusts to its new environment, but the lower soil horizons usually show little change unless they are mechanically disturbed.

Mycorhizae. The humus layer is the focal point of biological activity in forest soils. In addition to

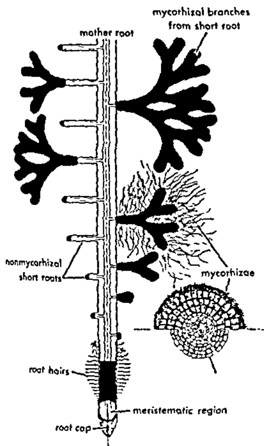


Fig 3. Development of mycorhizae on a pine root Solid black areas indicate absorbing surfaces The main axis is a mother root bearing a root cap and root hairs on the unsuberized surface behind the meristematic region. On the upper surface is shown the development of mycorhizal branches from short roots. On the lower surface are several nonmycorhizal short roots. The cross section above represents a mycorhizal root, below, an uninfected root. (From P. J Kramer, Plant and Soil Water Relationships, McGraw-Hill, 1949)

litter decomposing animals and microbes, it is inhabited by roots of forest vegetation. Portions of the root systems of many trees and other woody plants have modified roots called mycorhizae (Fig. 3) which are formed by association with certain fungi and play an important role in absorption of nutrients and moisture. Because the supply of moisture and available nutrient elements is generally favorable, many of the absorbing roots of trees and other forest vegetation are located in and immediately below the humus layer. This arrangement allows maximum utilization of the nutrient elements released by decomposition of the organic material Into this cycle are added the nutrient elements taken up from weathering rocks and minerals by roots occupying deeper soil layers. The depth of forest soils is governed by the depth of root penetration which may vary from a few inches to over 50 feet. Trees generally utilize a much larger volume of soil to maintain growth than do agricultural crops.

Classification. Forest soils may be classified according to a variety of criteria. The classification of the Soil Survey Division of the Bureau of Plant Industry, Soils, and Agricultural Engineering is sometimes used, but since this system is based on soil features important to agriculture, it is generally inadequate for forestry purposes Forest composition, minor vegetation, parent material, topography, depth to ground water, and growth and yield of forest stands all have been used in classifying forest soils Classifications based on forest productivity have also been found to be fairly successful. In such systems, growth of the aerial portion of trees is related to topography; soil texture, structure, aeration, and infiltration of water; moisture retention capacity, soil depth; and type and distribution of organic matter and nutrient supply, all of which also influence the growth of the subaerial portion of trees. These relationships are then used to predict potential productivity of understocked or deforested lands.

Hydrologic cycle relations. In addition to their role in producing cellulose, timber, and other forest products, forest soils perform an important regulatory function in the hydrologic cycle (see HYDROLOGY). Well-managed forested watersheds are often characterized by soils with a high water infiltration capacity Such soils are usually protected from raindrop impact, which causes surface sealing and promotes overland flow Excessive overland flow results in soil erosion Where falling rain or melting snow have maximum opportunity to penetrate into the soil, the natural drainage

[G.K.V.]

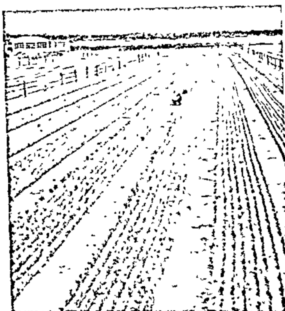
Bibliography · See FOREST AND FORESTRY.

Forest tree nursery management

An aspect of forestry concerned with the production of high-quality planting stock intended for artificial regeneration of forest stands Although the details of operation may vary somewhat with the region and species, there are a number of principles that are applicable to all situations.

Site factors. The selection of a permanent forest nursery site requires careful consideration of

inside as closely as possible. The nursery site should provide adequate air drainage to eliminate frost pockets; windswept areas and sites subject to periodic flooding are to be avoided. The area should be level if the soils are well drained. Otherwise, a uniform gentle slope in one direction is advantageous (Fig. 1).



The forest nursery. (*South Carolina State Commission of Forestry*)

The most desirable soils consist of 2–3 ft of uniform sandy loam underlain by a loam with good vertical drainage. Coarse-textured soils require more frequent irrigation which causes loss of nutrient elements by leaching. Fine-textured soils drain slowly and at times may be inadequately aerated. Stock lifting in such soils is often difficult, and the weed problem may be troublesome. In some climates fine-textured soils are also subject to frost heaving. Nursery soils should be free of stones, impervious or calcareous substrata, toxic agents, and any evidence of mottling. The natural level of available soil nutrients should be adequate for growth of common agricultural crops.

The site selected should be large enough to accommodate the requirements of the planting area and to allow for future expansion. A constant supply of good quality water is necessary. The nursery should be located close to a suitable labor force and be easily accessible to an adequate communication and road system.

Once selected, the nursery area is subdivided for efficient management. Space is allotted for buildings, roads, composting pits, watering systems, windbreaks, and fences. Trees and other vegetation are completely removed from the planting area, and the soil is plowed, disked, and harrowed until level and free from stones and other debris. If grading is necessary, the topsoil is scraped aside and replaced when the grading operation is completed. The seedbeds are laid out in uniform sizes. They are usually 4 ft wide and as long as the irrigation system will allow. Longer beds result in more efficient operation than short beds. The beds are prepared by hand, or by using bed-shaping equipment or self-powered rototillers. The surface of the bed should be smooth and free from coarse soil clods or organic material that might interfere with seed germination and seedling establishment.

Seeding and seedbed preparation. The source of seed used in a nursery is of extreme importance. If possible, only seed of local origin collected from trees of superior quality should be used. Seed from other geographic regions should be tested thoroughly. The sowing rate should produce about 25 to 60 conifer seedlings or from 10 to 25 hardwood seedlings per square foot, depending upon the species and soil fertility. Seed is sown broadcast or drilled in rows. When seed is broadcast, it is pressed into the soil with a light roller and covered with clean sand. Time of seeding varies with species; some require fall sowing, others are spring sown, and some do equally well at either season. After sowing, the beds are protected by mulching from frost, wind, and raindrop-splash erosion, excessive moisture loss, birds, and rodents. The mulch is usually removed before the seedlings appear above the ground level. Seedlings of shade-tolerant species should be shaded for the first season.

During seedling and transplant development, weed competition should be minimized by cultivation, hand removal, or by chemical control. In row-sown seedbeds, shallow cultivation may be practiced; in broadcast-sown beds, weeds are removed by hand. Either operation may be supplemented by chemical weed control using allyl alcohol, carbon disulfide, methyl bromide, and other compounds applied before or immediately after sowing to prevent weed seed from germinating. Mineral spirit derivatives from naphthenic petroleum may be sprayed directly on seedbeds of most coniferous species to kill growing weeds; however, these sprays kill or injure hardwoods.

In regions where the seedlings are kept in the nursery for more than one year, they may be root-pruned or transplanted at the end of the first or second season in the seedbed. Root-pruning does not replace transplanting but it promotes a fibrous compact root system and is considerably less expensive. Transplanting produces sturdier trees with better-developed root systems than are produced in the seedbed.

Field planting. When the seedlings or transplants have grown to a suitable size, they are lifted and prepared for field planting. If fall planting is scheduled, dormancy should be induced by decreasing the amount of water available to the plants in the late summer. This practice is known as "hardening off." In conifers, dormancy is indicated by the formation of the terminal bud and cessation of needle growth; in hardwoods, by leaf "shedding." The stock is lifted with a spade or by a mechanical tree-lifter. The trees are separated, graded, counted, and packed. The aim of grading is to insure stock free of insects and disease and to eliminate trees which do not conform to certain standards based on morphological characteristics such as average stem diameter or the ratio of top growth to root growth. More complex standards based on physiological characteristics correlated with drought and frost resistance and growth rate are also used. The

system for counting plants depends upon the scale of the operation. For very large shipments, the number of plants is based on inventory figures obtained during the previous growing season. For small shipments, a full count is required. The trees are bundled and packed in wet moss or shavings to form bales which are wrapped in heavy waterproof paper. Usually the bales of trees are shipped immediately, but it is sometimes necessary to store them. At temperatures of 34–40°F, trees may be stored safely for 4–6 weeks. Extended storage generally reduces survival and early growth rate in coniferous trees. Dormant hardwoods may be heeled in (*temporarily set out in soil beds*) over the winter without damage. From the initial lifting of nursery stock until it is field planted, extreme care must be taken to prevent exposure to drying winds or sunlight. The roots should be kept cool and moist at all times.

Water requirements. In most regions, supplemental irrigation is *essential for stock production.* The amount of water added by irrigation varies with climate, soil, species, and size of stock. The root zone should be *kept free of moisture stress* throughout the growing season, but excessive use of water should be avoided. The maximum amount added should *not greatly exceed the normal rainfall* for the same period. Thus, even in southern nurseries, about 1 in. per week during the growing season should suffice. In most overhead systems, this amount requires *about 8 hours of sprinkling* time. During germination and early seedling development, the upper soil zone should be kept moist by frequent light *additions of water.* When the seedlings have become established, heavier applications of water at longer intervals are more effective. On extremely *hot days, light sprinkling may* reduce heat injury.

Soil management. Because of the high plant population and complete removal *of the crop, along* with a considerable amount of adhering soil colloids, successful nursery operation demands a rigid program of soil management. Maintenance of favorable soil structure *requires* renewal of soil organic matter by plowing under leguminous or nonleguminous green manure crops or other plant or animal residues. Commercial *fertilizers are applied* prior to seeding the soiling crops, and when these crops are incorporated in the soil, the nutrient elements are slowly released from the decomposing plants. In soils with a low level of organic matter or where d.

...ese materials are usually composted before use. In southern nurseries, finely divided organic material such as sawdust is successfully used without prior composting, provided sufficient commercial fertilizer is added.

It is often necessary to supplement the mineral and organic soil amendments with additional mineral fertilizer applied dry or in solution. The beds should be well watered immediately *after fertilizer* application to prevent stock injury. The fertility status of the nursery soil should be determined annually by chemical analysis of soil and tissue samples. Minimum fertility levels have been established for some species by analysis of soils supporting productive stands of these species. Observations of test plots treated with different types and levels of fertilizers are useful in establishing optimum fertility levels. Soil reaction should be maintained at about pH 5.0–5.5 for conifers and about 5.5–6.5 for hardwoods. Some soils, particularly those treated with sulfuric acid, may require periodic light applications of dolomitic limestone to avoid deficiencies of calcium and magnesium since excessive acidity may render these two elements unavailable to the plants. Where necessary, calcareous soils are acidified by applying sulfur or sulfuric acid.

Pest control. Throughout its development in the nursery, the seedling is exposed to numerous pests such as birds, rodents, insects, and parasitic fungi. Birds and rodents are most serious during ... Bird dam-

rodents

poison-

ing. The larval forms of insects are most destructive, and effective controls involve the use of chlorinated hydrocarbon compounds applied as a dust or in suspension. Damping-off fungi commonly cause the most *extensive damage in many nurseries.* This disease is controlled by soil application ... aluminum sulfate, elemen-

also successful in many cases. Safe and effective application rates for herbicides, insecticides, and fungicides must be worked out in trial seedbeds. *See* FOREST AND FORESTRY, *see also* FOREST SEEDING AND PLANTING; PLANT DISEASE. [G.K.V.]

Bibliography: *See* FOREST AND FORESTRY.

Forest vegetation

A large area of dense growth composed essentially *of trees and* commonly containing bushes and *shrubs.* Trees are classified as coniferous if the

of it throughout the year. *In terms of density, closed canopy forests* are those in which the tree *crowns touch to form a* continuous *canopy; the trees in* scrub forests *are more* widely spaced with grass occupying the *intervening* area. Rain forests develop in especially humid *situations; their foliage* is dense and interlaced to form a closed community, and the trees support many *vines, creepers,* and *epiphytic* plants.

In the tropics the rainy lowlands bear tropical rain *forests of* evergreen broadleaf *trees; the up-* ... dry seasons or monsoonal climates

emidecid-

ison tropi-

cal forests are of scrub character and some are thorny.

Middle latitudes contain a variety of forests. In the warmer parts of the mid-latitudes the humid areas bear broadleaf forests, which in areas of poor soil are replaced or mixed with conifers. The winter rain areas of the subtropics support mediterranean scrub forests of broadleaf evergreen trees combined with much brushy growth The boreal or



Fig. 1. Tropical rainforest and riverside clearing in Peru (R G. Le Tourneau)

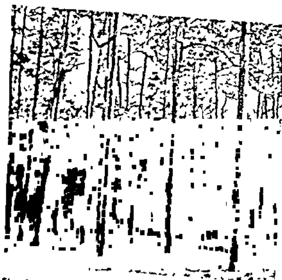


Fig 2. A ground view of long-needle evergreen pine forest, typical of the poor sandy soils of Atlantic and Gulf coastal plain in southeastern United States. (U.S. Forest Service)



Fig. 3. Mixed hardwood-conifer forest of birch, beech, maple, and hemlock in Wisconsin. (U.S. Forest Service)

northern coniferous forest, composed principally of firs, spruces, and larch, occupies the cooler subarctic parts of the mid-latitude continents.

The forests and their animal populations have been much altered by human occupance. However, the African tropical scrub forests are heavily populated by ungulates and their predators (see STEPPE). The animals of the boreal forests (beaver mink, martin, muskrat, sable, and others) supply some of the world's most valuable furs. *See* VEGETATION ZONES (WORLD); *see also* ECOLOGY; FOREST AND FORESTRY. [C.M D]

Forestry, farm

The application of the techniques of forestry to the management of farm woodlands. Basically these techniques deal with the producing, tending, harvesting, and marketing of forest-grown crops.

Farm woodlands aggregate approximately 165,-200,000 acres or 34% of the timber-producing land of the United States. These woodlands, owned by 3,400,000 individuals, contain 15% of the country's standing timber volume. The greater part, 91% of the farm woods area, is in the eastern half of the United States. Farm woods are usually small, the average size for the entire country being 49 acres; 56% of the individual units are under 30 acres in area. Timber volumes of farm forests are generally low, the average stand per acre being estimated at only 1900 board feet.

Farm woodlands are usually lower in productivity than forest lands in any other form of ownership, public or private. Seldom regarded as of ma-



Fig. 1. A heavily grazed wood lot of even-aged trees offers little soil protection, steadily declines in timber production, and harbors a minimum of wildlife. (U.S. Forest Service)

jor importance, farm woodlands in general receive little care or constructive management (Fig. 1). The timber produced on them has usually brought low prices, cutting methods have been destructive, and little attention has been given to reproduction and sustained yield. Destructive grazing by domestic livestock has been common; fires have been allowed to burn unchecked. The owners as a class have shown little interest in woodland management, primarily because they have been preoccupied with other farm operations, have been uninformed on the proper techniques of woodland management, and usually are at a disadvantage in marketing their forest products. Nevertheless, farm woodlands must contribute importantly to long-term production of the country's forest supplies and services because of the vast areas involved, their association with and contribution to operating farms, their proximity to population centers and markets, and their potential for growing timber.

Management. The management of farm woodlands involves the development and maintenance necessary to produce continuous yields of maximum quantity and quality (Fig. 2). Although timber is the primary crop, woodlands are also important in providing protection from wind and water damage, creating favorable habitats for wildlife, and furnishing areas for recreational use. The broad aims in managing farm woods are identical with those in managing forest lands under other form of ownership. Likewise, the techniques of management are similar, though the problems to be solved in dealing with farm woods usually are simpler than those associated with handling the larger forest areas. The practices involved are those of mapping woodlands and determining the volumes of standing timber; applying of silvicultural techniques of reproducing natural stands; reforesting to start new woodlands; caring for woodlands through their life cycles; protecting woodlands from destructive agencies, principally domestic livestock, fire, injurious insects, and diseases; harvesting of the tree crop so that the residual stand is undamaged and left with an adequate stock for producing future yields; and marketing procedures which result in the best interests of the owner.

Operations. The solution to the farm forestry problem is engaging the attention of both public and private agencies in all parts of the United States. Federal aid is usually supplied through cooperative laws of which the Clarke-McNary Law of 1924 and the Cooperative Forest Practice Act of 1950 are of outstanding importance. Under these laws the Federal government annually allocates money to state colleges of agriculture or state de-



Fig. 2. (a) Sustained-yield [illegible] woodland may increase the farm income considerably. (b) Demonstration of selective cutting in a farm wood lot. (USDA)

partments that deal with forestry. To the Federal allocations the state agencies contribute their required shares and then administer the expenditure of the total sum which goes toward employing trained personnel to give advice and field assistance, and to some extent toward providing planting stock. Many states also have enacted laws to meet local conditions. Most of the technical aid is provided from the colleges of agriculture through their experiment station and extension services and from the state departments of forestry and conservation. This aid is supplied through personal visits for education and service, letters, publications, radio, and other media. In some sections, valuable advice is likewise furnished by some of the wood-using industries. These companies have a direct interest in insuring a continuous future supply of raw materials for their paper mills or other wood-using plants. Many industries have their own foresters who are available to give service in all aspects of forestry to the wood producer.

Progress is being made in getting farm owners to apply scientific forestry techniques to the handling of their woods. Recently in some states the total annual value of farm forest products ranks near the top in agricultural enterprises. However, before the problem can be solved entirely each owner must recognize that his woodland is an important part of his over-all farm operation, and that it can be managed so as to provide a continuous return of products, services, and income. *See* AGRICULTURE; SILVICULTURE. [C.H.C.]

Bibliography: *See* FOREST AND FORESTRY.

Forging

The plastic deformation of metals, generally at elevated temperatures, using compressive forces exerted through a die.

General characteristics. Except where otherwise stated, forging deforms the metal at temperatures and strain rates such that no strain-hardening results; that is, the deformation is carried out above the metal's recrystallization temperature. As a result of the hot-working, nonmetallic inclusions are elongated in the direction of working, producing a fiberlike structure called grain flow, flow lines, forging fiber, or fiber structure. This structure causes the forging to be anisotropic with respect to mechanical properties [illegible] to ducti[illegible] greater [illegible] degree of anisotropy developed depends on the amount of reduction effected during forging.

Forging operations are generally carried out at elevated temperatures in ordinary atmospheres, hence the surface oxidizes to an extent depending on the characteristics of the particular metal. The usual results of this oxidation are formation of scale (oxide coating), toughened surface, and limited dimensional control. On steel parts decarburization and internal oxidation may also occur with a detrimental effect on fatigue properties.

Preparation of work. Raw material used for forging depends primarily on the type of process; it is generally in the form of a bar or billet. In most instances the shape of the part precludes the forging being done in a single operation. This restriction is due primarily to the frictional effect and the cooling action of the die on the hot metal. As a result, a series of preforming operations is required which moves the metal in steps toward the finished shape. Thus the production of a forging normally consists of one or more of the following operations:

1. Drawing—spreading or thinning action reduces thickness and increases length.
2. Fullering—similar to drawing; results in an elongated section between two heavier sections.
3. Edging or gathering—a preliminary step wherein metal is moved or distributed to facilitate filling of die cavity in subsequent operations.
4. Bending—a final operation or a preliminary step in preshaping the blank to agree with impressions in finishing die.
5. Upsetting—shortening and thickening action due to axial force on bar.
6. Blocking—confined flow which distributes metal to approximate shape of finished part.
7. Finishing—confined flow which shapes metal to final form.
8. Trimming—removal of fin or flash which occurs in finishing.
9. Punching—displacement of metal by plastic flow to form a hole.

Classification of processes. Forging processes are usually classified by the type of equipment used to form the piece. However, some operations may be performed on more than one type of equipment. The usual classifications are hammer or smith forging, drop forging, press forging, upset or machine forging, and roll forging. Swaging and cold heading are related to forging but are not necessarily hot-working processes.

Hammer (smith) forging. The application of mechanical power hammers to the blacksmith type of operation constitutes hammer or smith forging. The anvil of a forging hammer holds the lower die (usually flat, but may be V-shaped or rounded) and is not necessarily an integral part of the frame. The operation consists of repeated blows of the hammer transmitted through the upper die to the work. The reduction produced by each blow and the pressure required are under the control of the operator. Although the process has a low production rate and is limited to simple shapes and large tolerances, it has the advantages of wide range of size and low tooling cost. As the pressures are limited on mechanical power hammers, large open-die forgings are produced on hydraulic presses.

Drop forging. Forgings produced on drop hammers using closed dies are drop forgings. Two basic types of equipment are used, the board drop hammer and the steam drop hammer. On each the anvil

holding the lower die is an integral part of the frame of the machine, and the impact blow is delivered by a weight, or ram, which holds the upper die. In the board drop hammer, the ram is attached to wooden boards which slide between two rollers. The ram is raised by friction between the rolls and the boards and, upon reaching the desired height, is allowed to fall freely, the forging force (up to 1000 lbs) thus being developed by gravity. In the steam drop hammer, the ram is attached to a steel piston which is raised or lowered by a double-acting steam cylinder. The forging force (up to 50,000 lbs) is a result of the combined effects of gravity and steam pressure. In both types of hammer the heated metal is forced to fill a die cavity by repeated impact blows, the force of which is under the control of the operator. Drop hammers may also be used to produce open-die forgings and for sheet metal forming.

To maintain dimensional accuracy in drop forgings, the two dies are carefully aligned and come together in the finishing operation. However, as it is not feasible to provide the exact amount of metal to fill the die cavity, provision is made in the dies to allow a small excess amount to be forced out of the main die cavity into a small cavity called a flash gutter. The resulting flash (Fig. 1) or fin extending from the forging at the parting line is then trimmed off.

Impact forging. To increase the effectiveness of the forging blows, the impacter forging machine was developed. Two horizontally mounted rams are driven toward each other at high velocity by air pressure. The stock is positioned at the plane of impact of the rams (Fig. 2). Thus the energy of the rams is almost entirely absorbed by the metal and very little energy is lost by transmission to the machine foundation.

Press forging. Some parts may be produced by a series of steps each of which is completed by a single application of the forging force, in contrast to the multiple blows of the drop hammer. This may be done in a mechanical or hydraulic press. The method is particularly adaptable to symmetrical parts requiring few operations. For extremely large forgings, the hydraulic press is the only equipment which can provide the necessary forces (up to 50,000 tons). In addition to the usual vertical rams, presses may be equipped with horizontally operating rams permitting simultaneous forging in several directions.

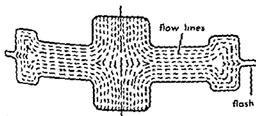


Fig. 1. Typical drop forging cross section showing flow lines and flash.

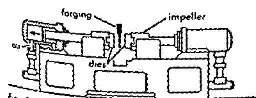


Fig. 2. Impacter forging machine. (Chambersburg Engineering Company)

Upset (machine) forging. This process consists of upsetting a heated bar in a successive series of operations which increase the diameter, shorten the length, and cause the stock to fill a die cavity. The operations are carried out in a horizontal double-acting press called a forging machine or upsetter. Split dies, which move transverse to the axis of the bar, grip the stock and can also bend, shear, punch or trim the stock. The punch or heading tool may also perform several operations other than simple upsetting. Thus the forging action may be accomplished in two directions. Usually a multiple die setup is used in one machine so that the series of forging operations can be performed quickly without the necessity of reheating the metal between operations. The process is more automatic than other forging methods, thus labor costs are reduced. The upset forging method is generally limited to cylindrically shaped parts of relatively small size (maximum bar size of about 9 in.) and to certain dimensional ratios.

Roll forging. Preforming of blanks prior to press or hammer forging is often accomplished with forging rolls. These are semicylindrical rolls containing several grooves, permitting a sequence of operations. The contour of the grooves determines the shape of the final product. The process is particularly adaptable to drawing metal into long, slender bars, with either straight or tapered sections, such as automotive rear axle shafts.

Swaging. While not ordinarily classified as a forging process, swaging (or rotary swaging) is similar to forging in that plastic deformation is produced by compressive forces exerted through hammer blows. The blows are obtained by dies located in a slotted, rotating spindle. Centrifugal force causes rounded blocks, which are in contact with the dies, to strike a series of axially mounted rollers contained in a cage surrounding the spindle (Fig. 3). This action causes the die halves to close and to compress the metal being swaged. Centrifugal force opens the dies after the backing blocks pass the rollers, permitting the stock to be rotated or moved back and forth.

The process may be carried out on hot or cold bar stock or tubing to reduce the diameter over part or all of the length. Tapered or stepped sections are readily produced; and by the use of mandrels, internal threads may be produced in tubing. Also, by machining a tubing blank to predetermined outside shape and then swaging the outside to a

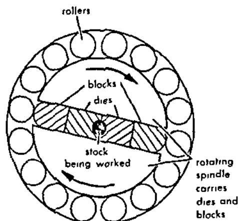


Fig. 3. Swaging machine—principles of operation.

straight, uniform diameter, the outside contour is transferred to the inside. Cold swaging results in excellent surface finish, accurate dimension, and higher strength due to strain-hardening.

Cold heading. This process is similar to upset forging but it is done on cold metal and is limited to relatively small parts (1-in. stock) due to the strain-hardening that occurrs in any cold-working operation. The process originated for heading rivets, bolts, and screws, but it is considerably more versatile. The machine is much like a forging machine in principle except that the operation is more automatic. Wire stock is automatically fed through a die, the unsupported section being upset by the punch and assuming a controllable shape. The pieces are cut to length and ejected from the machine. A production rate of as much as 800 pieces/min may be attained.

Design consideration. The design of forgings is more restricted than that of castings, due to the inability of solid metal to flow under pressure to the same degree as liquid metal flows in filling a mold cavity. To facilitate the flow of metal in the die, the following general design principles should be followed wherever possible: (1) the parting line should be in one plane, flat, and located near the centerline of the forging; (2) radii and fillets should be as large as possible; (3) pockets, recesses, small holes, and sharp corners should be avoided; (4) ribs should be low and wide; (5) sufficient draft and generous tolerances should be provided to permit removal of part from die. *See* METAL FORMING.

[R.L.F.]

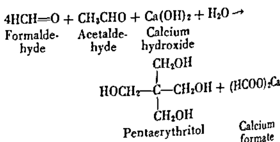
Bibliography: P. E. Kyle, *The Closed Die Forging Process*, 1954; W. Naujoks and D. C. Fabel, *Forging Handbook*, 1939.

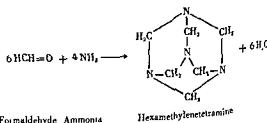
Formaldehyde

[illegible]

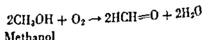
Therefore, it is produced and marketed as an aqueous solution (usually 37% formaldehyde by weight), sometimes known as Formalin. It is also sold as the solid hydrated polymer known as paraformaldehyde or paraform. Production of formaldehyde is now at the rate of about 500,000,000 lb per year calculated on an anhydrous basis.

Uses. Formaldehyde is used principally to produce synthetic resins by reaction with phenols, urea, and melamine. This use accounts for about 75% of the total production. Approximately 15% is used in the manufacture of textiles, dyes, drugs, paper, leather, photographic materials, embalming agents, disinfectants, and insecticides. Increasing amounts are used for the industrial production of pentaerythritol and hexamethylenetetramine by the following reactions:



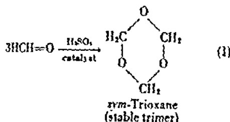


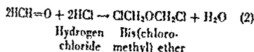
Production. Most of the formaldehyde is manufactured by the oxidation of methanol with air over a metal catalyst in the temperature range of 450–650°C. Catalysts are either silver, copper, or an iron-molybdenum mixture.



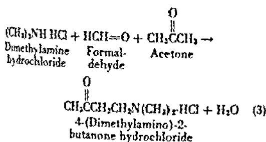
Increasing quantities of formaldehyde are produced by the partial oxidation of hydrocarbons from natural gas. Air is used as the source of oxygen, excess hydrocarbon to minimize complete oxidation, and steam to quench the reaction. The yields of formaldehyde are low, but the raw materials are plentiful and low in cost. Since other products (such as alcohols, acids, ketones, and other aldehydes) are formed, the purification of formaldehyde produced by this process may involve fractionation, liquid-liquid extraction, azeotropic distillation, and extractive distillation.

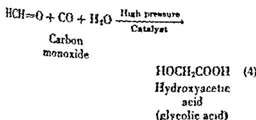
Properties. Formaldehyde undergoes many of the general reactions of aldehydes (*see* ALDEHYDE). Some interesting reactions which are somewhat specific for formaldehyde are the following:

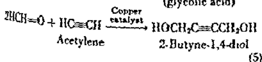




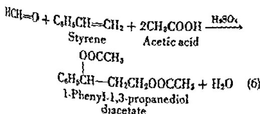
The Mannich reaction (3) involves the reaction of an amine hydrochloride with formaldehyde and a ketone.







The Prins reaction (6) involves the reaction of formaldehyde with olefinic aromatic hydrocarbons in the presence of an organic acid and a mineral acid catalyst.



The reaction of formaldehyde with phenols, urea, and melamine to form resins proceeds first with formation of methylol ($-CH_2OH$) derivatives followed by intermolecular dehydration to produce methylene (CH_2) linkages. This reaction is catalyzed by either alkali or acid, and the extent of resinification is controlled by temperature, reaction time, and concentration. *See* PHENOL-FORMALDEHYDE RESIN. [H.A.S.]

Formic acid

A colorless, pungent liquid having melting point 8.4°C and formula HCOOH. Formic acid is miscible with water, alcohol, ether, and glycerol. It is toxic, and causes painful skin wounds. It is a constituent of the stings of ants, stinging caterpillars, and stinging nettles.

Sodium formate is produced by heating carbon monoxide with sodium hydroxide under pressure at about 210°C. Treatment of sodium formate with sodium acid sulfate or dilute sulfuric acid liberates free formic acid.

Formic acid is stronger than acetic acid. With carbonates, oxides, and hydroxides of alkali or alkaline-earth metals, it gives formates (salts); with alcohols it forms esters (formates). *See* ESTER

It is used as a reducing agent for metallic ions and dyes, in dehairing and tanning, as a latex coagulant, in the preparation of a large number of esters, as a source of allyl alcohol via its reaction with glycerol, and in the preparation of oxalic acid (by heating of sodium formate) With concentrated sulfuric acid, it decomposes, liberating carbon monoxide. *See* CARBOXYLIC ACID; OXALIC ACID. [E.B.R.]

Fossil

The remains and traces, such as footprints and trails, of organisms preserved from the geologic past The term fossil was originally used for a wide range of curios, both organic and mineral, found in rocks, but it came to be restricted to organic remains about the close of the seventeenth century. The term fossil is still used as an adjective, and in a figurative sense, to indicate that nonorganic structures in the rocks were formed in the geologic past. Examples of nonorganic structures are fossil mudcracks, fossil rain imprints, fossil ripple marks, and even fossil fuels

Nature of fossils. Several different modes of preservation are known. Generally only the hard parts such as bones and shells, have been preserved, but in some instances even the soft tissues are preserved almost intact

Unaltered remains. Extinct mammals such as the woolly mammoth have been found in the ice or frozen ground of the Arctic where they remained in cold storage for some thousands of years. The Beresovka mammoth, found in eastern Siberia in 1901 was virtually complete, with clotted blood in the chest and unswallowed food in the mouth. Part of the flesh was still red and was eagerly devoured by the collectors' dog team. In Poland, entire carcasses of the extinct woolly rhinoceros are preserved in oil seeps. Such preservation of animal tissues is limited to late glacial and postglacial deposits. Fossil wood is more durable, however, and is found virtually unaltered in rocks as old as the Cretaceous.

Petrified fossils. Some organic objects are turned to stone. This usually involves only the hard parts

such as bones, shells, and wood. Petrifaction takes one of three forms:

1. Permineralization involves only the addition of mineral matter from underground solutions which fills the voids (mostly microscopic) in solid structures such as bone, shell, or wood.

2. Replacement involves the substitution of some other mineral, from underground solutions, for the original material. This is commonly accomplished molecule by molecule so that even microscopic structures are preserved while the object is literally turned to stone. The fossil wood shown in Fig. 1 illustrates both replacement and permineralization. The original cellulose of the cell walls has been replaced by quartz and the voids of the empty cells have been filled by quartz.

3. Distillation concerns organic tissues, such as cellulose and flesh, that are largely made of compounds of carbon, hydrogen, and oxygen. After burial, and in the absence of oxygen, such compounds undergo destructive distillation that liberates carbon dioxide and water until only free carbon remains. This carbon forms a black film in the rock, a carbon copy of the animal or plant. Fossil leaves are commonly preserved in this way and, more rarely, the outline of the fleshy body is preserved surrounding the skeleton of a fossil animal (Fig. 2). Even the remains of completely soft-bodied animals are thus preserved in the Burgess black shale of Middle Cambrian age near Field, British Columbia.

Molds, casts, and imprints. These form another category of fossils. They are formed when the sediment has solidified about an organic object and the latter is subsequently dissolved, leaving a hole in the rock. This is a mold. By pressing a plastic substance such as dental wax into the mold an artificial cast of the object is obtained. Deposition of mineral matter from underground solutions may also fill the hole, producing a natural cast. Fossils which preserve the form but not the substance of the originals are called pseudomorphs. Molds of thin objects such as leaves are commonly called imprints. Figure 3 illustrates the imprint of a fern leaf which has undergone distillation. The carbon

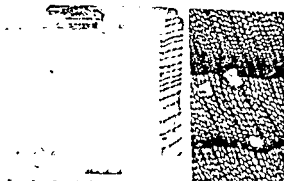


Fig. 1. Petrified wood from the Fossil Forest of the Yellowstone; a thin slice is enlarged to show the cells in three growth rings.

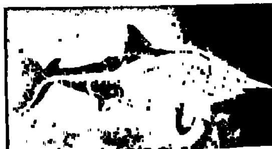


Fig. 2. A marine reptile, *Ichthyosaurus*, from the Lower Jurassic black shale at Holzmaden, Germany. (*Courtesy of the American Museum of Natural History*)



Fig. 3. Fern leaf from the Coal Measures.

film faithfully records the shape of the leaf, but even if the carbon were removed it could be recognized from the imprint in the rock.

Tracks, trails, and burrows. These traces constitute a fourth category of fossils. Although involving no remnant of the organism, they are structures made by a living animal and, as such, may give evidence of its size, shape, and habits. They tell, for example, whether a dinosaur was bipedal or quadrupedal, whether it was slender and agile or ponderous and slow, and whether it ran like an ostrich or leapt like a kangaroo (Fig. 4).

Even worm tubes in an ancient sandstone prove that living organisms were present and may give clues as to the local environment under which the rocks were formed.

Coprolites. These fossils are petrified excrement. Undigested particles of food preserved in such fossils may indicate the feeding habits of animals long since extinct. Plant fragments in the dung associated with the extinct ground sloths, for example, have demonstrated the kind of vegetation they preferred.



Fig. 4 Tracks of a small bipedal dinosaur in Triassic sandstone from Turners Falls, Mass.

Uses of fossils. The fossil record is of major importance to the geologist. Several aspects of this record are considered.

Correlation. Fossils permit correlation of sedimentary rocks of equivalent age the world over. By this means a synthesis of local details into broad regional interpretation is possible.

The use of fossils in correlation was discovered accidentally by W. Smith just before 1800, as a result of field work in the richly fossiliferous Jurassic rocks of southern England. In that region a long sequence of formations (alternating limestones and shales) lies gently tilted and crops out in bands that are easily traced across the country. After extensive collecting, Smith perceived that each of these formations carries a unique assemblage of fossils by which it could be recognized in any outcrop. This insight was soon used by other geologists and found to be widely applicable. Fossils serve as a guide or index to any fossiliferous rock and characterize the deposits of a limited part of geologic time. See INDEX FOSSIL.

Smith had no idea why this should be so. It was an empirical discovery based on tedious and detailed observations and rested on no theory. Darwin's *Origin of Species*, published in 1859, provided the explanation. Life has been steadily evolving throughout geologic time and each genus and species had a relatively short duration; thus the faunas entombed in successive ages must necessarily differ each from the next. See STRATIGRAPHY.

Chronology. Fossils date the rocks in terms of geologic time. Once the sequence of faunas has been worked out in a long succession of undisturbed sedimentary deposits, it is possible to determine the relative ages of the fossils by the law of superposition, the oldest formation being the lowest and each in turn being younger than the one on which it rests.

Of course no single region contains a complete record of all geologic time, and if it did the succession of strata would be so thick that the lower part would be deeply buried and inaccessible. The fact that part of the record is preserved in one region and other parts in other regions makes it necessary to correlate only enough partially overlapping sequences to establish a composite record for all of geologic time. In this way the Geologic Column has been built up by the collaboration of geologists throughout the world and the corresponding Geologic Time Scale has been established in terms of which events can be placed in proper chronologic order. If the stratified rocks are a manuscript in stone, fossils supply the pagination. No other principle has had such far reaching influence in geology. See GEOLOGIC TIME SCALE; GEOLOGY.

Past environments. Fossils indicate the local environment under which many of the stratified deposits were laid down. The presence of corals, echinoderms, brachiopods, cephalopods, and other such exclusively marine organisms proves that the enclosing sediments were laid down on the sea floor, even though they are now far inland and even high in mountain crests. On the other hand the presence of abundant land animals or of land plants, and of stumps in their place of growth prove deposition to have taken place on land. More specific details about the environment may be indicated by certain fossils, especially when their study is combined with that of the sediments. Coral reefs, for example, will signify a warm and shallow sea floor. Other associations may indicate shallow mud flats, and still others broad shallow sand flats. The presence of exclusively pelagic organisms in dark shale or limestone will suggest deeper water and foul stagnant bottoms on which benthonic organisms could not survive. Certain types of vegetation will suggest a swampy lowland, whereas buried stumps and standing tree trunks in sandstone indicate rapid deposition and quick burial on a river flood plain. These are only samples of the use of fossils in judging the environment of deposition of rocks of the past. See PALEOECOLOGY; SEDIMENTATION (GEOLOGY).

Geography of past ages. When the local environment of a sufficient number of areas has been worked out and all the deposits of a given age are correlated it becomes possible to construct a map showing the distribution of lands and seas of a particular geologic age for a limited region, a whole continent, or even for the whole earth.

It is known, for example, that for a time during the middle of the Cretaceous Period, North America was divided into two land masses separated by a vast inland sea that stretched from the Gulf of

Mexico to the Arctic Ocean. *See* Paleogeography.

Record of life. The history of life on earth and evidence of organic evolution are provided in the fossil record. Because soft-bodied organisms are rarely preserved there are numerous groups of animals and plants whose history is virtually unknown, but for those possessing hard parts (skeletons, shells, or woody tissue) a vast amount of knowledge is now at hand and is still growing as more and more fossils are discovered.

In regions where deposition was nearly continuous for considerable spans of time, fossils from successive horizons show an orderly and progressive change in the morphology of many different groups of animals or plants. The evolution of the horse from a tiny ancestor with four toes on the front feet and three on the hind feet during Cenozoic time is one example of the many phylogenies that are clearly documented by fossils.

Connecting links showing a common ancestry of stocks that have now diverged widely are not uncommon. A classic example is *Archeopteryx*, the Jurassic bird whose reptilelike teeth, claws, and tail clearly indicate that birds evolved from reptiles. The theriodont reptiles of Triassic time likewise record almost every step in the evolution from reptile to mammal. In smaller details the phylogeny of many smaller taxa is known from suites of fossils gathered from successive horizons in the rocks. *See* Evolution, organic; Paleobotany; Paleontology.

Ancient climatic conditions. Many groups of modern animals and plants are adapted to rather restricted climatic environments; the musk ox and reindeer, for example, to the Arctic tundra; the crocodile, great tortoises, and palm trees to the tropics and the subtropics; and hardwood forests to the Temperate Zone. Wherever such organisms are preserved in sedimentary deposits far from their present range the indications are strong that the climate has changed.

Caution must be used, however, because some groups are highly adaptive and individual species may have adjusted themselves to conditions far from the norm. The woolly mammoth and the woolly rhinoceros, for example, were Arctic adaptations during Pleistocene time of groups otherwise limited to the tropics, but a whole fauna or flora is unlikely to show such abnormal specialization. The rich flora in the Cretaceous rocks of southwest Greenland, for example, records a forest of deciduous trees normal to the Temperate Zone with some species such as the breadfruit tree now confined to the tropics. Surely this barren, treeless land had a much milder climate in Cretaceous time. Fossil plants have also been found in the rocks jutting out of ice-bound Antarctica. During the ice ages of the Pleistocene epoch the reindeer spread southward across Europe and was the chief game animal of prehistoric man. In America at the same time the musk ox migrated southward into New York and into Arkansas and Utah. For this period the fossil evidence of cold climate is abundantly confirmed by glacial deposits. During middle Cenozoic time Nevada was occupied by a flora composed of genera now restricted to the low, humid, coastal plain of Louisiana and eastern Texas; it could not therefore have been arid as it is today.

The evidence of past climates is clear for Cenozoic time when the plants and invertebrate animals were mostly of genera and even of species which are still living; the evidence of climates is less and less obvious, however, in earlier periods when the investigator must judge by extinct genera or larger taxa of organisms long since extinct. *See* Paleoclimatology. [C.O.D.]

Bibliography: C. O. Dunbar, *Historical Geology*, 2d ed., 1959; C. L. Fenton and M. A. Fenton, *The Fossil Book*, 1958; G. G. Simpson, *Life of the Past*, 1953.

Fossil fuel

A combustible solid, liquid, or gas formed in past geologic ages. Principal fossil fuels are coals (solids which are largely organic), oil shales (solids which are largely inorganic), petroleum (largely liquid hydrocarbons), tar sands, and natural gas (largely gaseous hydrocarbons). Each of these varies widely in composition and properties. No single theory seems to account for the formation of all types of coal or even for all deposits of a single type, and the same thing is true for all other fossil fuels. The ages of deposits may vary by many millions of years, but they all contain carbon and hydrogen. Peats resemble some fossil fuels but are relatively young. *See* Coal; Fuel; Natural gas; Oil sand; Oil shale; Petroleum. [E.A.Y.]

Fossil man

All prehistoric skeletal remains of man which are archeologically earlier than Neolithic (necessarily an imprecise limit), regardless of degree of mineralization or fossilization of the bone, and regardless of whether or not the remains may be classed as *Homo sapiens*, the type of modern man.

Discoveries began in the early years of the nineteenth century, although their meaning and antiquity were not recognized before the finding of the Neanderthal man in 1856. Remains have come principally from Europe and adjacent Asia, Africa, North China, and Java. Because of the apparent late entry of man into the New World, American Indian remains are all relatively recent, and recognizable as *Homo sapiens*.

Dating fossil man. Old World fossils of man (Hominidae) are confined to the Pleistocene, with the possible exception of the lower Pliocene *Oreopithecus*. *See* Primates (fossil). Dating within the Pleistocene is accomplished not only by conventional paleontology, that is, by faunal and geological associations, but also by association with human implements and by physical and chemical dating methods, such as radiocarbon and fluorine.

Carbon-14, a radioactive isotope with a half-life of 5568 years, forms a fairly constant percentage of the carbon atmosphere and hence of living matter. Its rate of decay allows an estimate in actual years of the time elapsed since death (charcoal, wood, shell, and vegetable matter are especially suitable). The range of this method is approximately 70,000 years back from the present. See RADIOCARBON DATING.

Fluorine from ground water is captured by the mineral matter of bone, accumulating slowly over long periods. The percentage of fluorine in a fossil bone may be assayed from microsamples. Although it does not give an age in years, this test provides good evidence of whether the bone in question is of the same age as other fossils in a deposit or is an intrusion, perhaps by burial. A related test uses radioactive uranium, in the form U_3O_4.

Other tests utilize the decline of nitrogen or phosphates, or the fluorine-phosphate ratio, F/P_2O_5, as an indication of relative age.

Lower Pleistocene. The remains of *Australopithecus* are the earliest and most primitive fossil hominids of the Pleistocene (see table). They are dated by association with an essentially Villefranchian fauna, and are believed to fall in the Kageran Kamasian Interpluvial and the Kamasian Pluvial, presumably corresponding with the First, or Günz-Mindel Interglacial and the Second Glacial of the Northern Hemisphere. The Australopithecinae are set off as a different subfamily from "true" men, for which a second subfamily, Homininae, is provided. The latter are distinguished by smaller molar teeth and increasingly large brain size. Its members are called euhominids.

The earliest euhominid fossils derive from the latter part of the early or the beginning of the middle Pleistocene (Second Glacial) and are mostly fragmentary. *Telanthropus*, known from three jaw fragments, comes from Swartkrans, one of the South African australopithecine sites. Although believed to be a small individual of *Australopithecus* (not a euhominid) by some writers, it differs from the latter in having smaller teeth and a better definition of the inferior nasal margin. Heidelberg man is known from a complete mandible, of primitive human form but with advanced teeth. Kanam man consists of a chin fragment from East Africa, of surprisingly modern appearance in the apparent prominence of the chin itself. The extremely early date is questioned, and the specimen also shows signs of a tumor which obscures its actual shape somewhat. Java man, found in two successive geological horizons, is the only early form known from good material. The brain, expressed in a cranial capacity of about 850 cm^3, was slightly above that of *Australopithecus*, but skeleton, skull, and teeth were characterized by all essential human features, except that there was present a diastema or precanine gap in the upper tooth row. See HEIDELBERG MAN; JAVA MAN.

Thus, the end of the lower Pleistocene saw euhominids, or "true men," present in Europe, Africa, and Asia, who not only overlapped with the australopithecines chronologically but also differed in type among themselves enough to suggest that evolution and divergence had been taking place for some time.

Middle Pleistocene. During this time (corresponding to the Mindel or Second Glacial, and Mindel-Riss Interglacial of Europe) there also is evidence of Java man, associated with the Trinil fauna of Java. Peking man, a slightly more advanced but closely related form, is dated as Second Interglacial at Choukoutien, near Peiping. Characterized by an average cranial capacity of about 1150 cm^3 and a smaller dentition than Java man's, his cave remains include evidences of cannibalism, fire, bone tools, and a stone industry. An apparently related human stock inhabited Northwest Africa from the beginning of the middle Pleistocene, known only from lower jaws but sharing specific traits of these and of teeth with Java and Peking man. The earliest, from Ternifine, near Oran, were associated with an early Acheulian stone culture. Three similar but less primitive jaws from the Atlantic coast of Morocco (Rabat, Temara, and Casablanca) date considerably later (Third Glacial or early Third Interglacial) and may indicate the survival and evolution of this stock in the same region. See PEKING MAN; TERNIFINE MAN.

This period in Europe is represented by the Steinheim skull and the Swanscombe skull vault; both are late Second Interglacial. Although they are of small to moderate brain size (about 1180–

Generalized correlation of human fossils with geologic stratigraphy*

Types of human fossils in Eurasia and Africa: *Homo* (non-sapiens) / *Pithecanthropus*	Types of human fossils in Eurasia and Africa: *Homo sapiens*	Glacial/Interglacial stages (Alpine terminology)	
		Postglacial	
	Cro-Magnon	Würm Glacial	Main Würm
Solo, Rhodesian, Neanderthal			Göttweig Interstadial
Neanderthal			Early Würm
	Fontéchevade, Kanjera	R/W Interglacial	
		Riss Glacial	
Peking	Swanscombe, Steinheim	M/R Interglacial	
Ternifine, Java		Mindel Glacial	
Heidelberg, Australopithecinae		G/M Interglacial	
		Günz Glacial	
		Pre-Günz stages	

* Vertical scale not calibrated to absolute time

1250 cm³, both being apparently female), they resemble modern man (*Homo sapiens*) in conformation of the vault, except for greater thickness, and are distinct in type and special features from the Java-Peking strain. No mandibles are definitely known from Europe for the middle Pleistocene, so that direct comparisons cannot be made with the northwest African fossils, but teeth and other evidence, including a single parietal bone of the skull from Ternifine, indicate that the two populations were distinct. The frontal bone of the Swanscombe vault is missing, but the Steinheim forehead is marked by heavily developed brow ridges. Resemblance to modern man in these crania is therefore only partial. *See* STEINHEIM MAN; SWANSCOMBE MAN.

Fragments of several skulls were found by L. Leakey at Kanjera, Kenya, in 1932, in associa- [illegible]

exact age is thus an important matter; it is possible that the fauna is somewhat later than has been assumed.

Upper Pleistocene. Existence of *Homo sapiens* during late Pleistocene time is attested by two cranial fragments from the Fontéchevade cave in France, which are reliably dated to the Third Interglacial. Neanderthal remains occur from the latter part of this stage in Europe, and are more numerous from the early part of the Fourth or Würm Glacial in Europe, North Africa, and western Asia. Equally late are the Rhodesian men of southern Africa and the Solo men of Java. All these types have a massive cranial structure, low skull, and large brow ridges. This is in contrast with the narrower and relatively higher skull vault of *Homo sapiens*, and with his reduced brow ridges and retracted facial skeleton, which exhibits a canine fossa, or hollowing of the maxillary region of the cheekbone, reflecting deflation of the sinus area. Solo man in particular remained primitive and small-brained, whereas Neanderthal man attained a brain volume equal to that of modern man. *See* NEANDERTHAL MAN; RHODESIAN MAN; SOLO MAN.

Homo sapiens entered Europe during a milder phase of the Fourth Glacial, the Göttweig Interstadial, dated by carbon-14 measurements as lasting from about 42,000 B.C. to 28,000 B.C. Invading users of upper Paleolithic flake-blade tools dispossessed the Neanderthals of Europe toward the middle of this phase. All were of Caucasoid racial type, though with minor distinctions, possibly tribal in nature. The earliest, represented by the Combe-Capelle and Lautsch men, were especially long-headed and narrow-faced; the Cro-Magnon type, associated with the Aurignacian stone-working tradition, was long headed but broad-faced, with low orbits. *See* CRO-MAGNON MAN.

Evidence outside of Europe is poor but suggests that *Homo sapiens* occupied most of the Old World at about the same time and that all other forms of men were either absorbed or became extinct. American Indians may have entered the New World near the beginning of the same stage, but the oldest known remains are probably those from Midland, Texas, about 15,000 B.C. In Australia, the Keilor skull from near Melbourne, of the type of the existing aboriginals, has been dated to approximately 7000 B.C.

Classification of human fossils. The fossil evidence has been organized in several different ways, with similar results. The temporal sequence given above suggests overlapping stages of development, with a primitive euhominid stage arising from an australopithecine stage in the lower Pleistocene and existing through the middle Pleistocene. This sequence was succeeded by several lines exhibiting larger brain size during the upper Pleistocene. One of these, *Homo sapiens*, of more progressive cranial form, was the main or sole survivor in late Pleistocene times. F. Weidenreich divided the euhominid types into three subfamilies based on morphological criteria, as follows: (1) Archanthropinae including Java and Peking man as well as the late but primitive Solo man, also *Meganthropus* of Java, which he believed to be a forerunner of Java man (*Pithecanthropus*); (2) Paleoanthropinae including the large-brained but cranially unreduced types of the upper Pleistocene, the Rhodesian and the Neanderthal forms; and (3) Neoanthropinae including modern racial types and their upper Paleolithic precursors, *Homo sapiens* in the usual sense.

W. E. L. Clark would reduce all hominid distinctions to three genera: *Australopithecus* for the australopithecines, *Pithecanthropus* for the Java-Peking group, and *Homo* for all more advanced forms. *Pithecanthropus* has the species *erectus* (Java) and *pekinensis* (Peking). Because of incompleteness of material, Clark recognizes only two species of *Homo: sapiens* for modern man and the pre-Neanderthal men of the upper Pleistocene, and *neanderthalensis* for the specialized Neanderthal men of the Fourth Glacial. The three genera are regarded as successive developmental stages, essentially corresponding to lower, middle, and upper Pleistocene, with a minimum of chronological overlapping.

Beyond the above, there is no acceptable formal classification of the human fossils, fragmentary as they are. Older writers gave taxonomic names to each new find with little regard to taxonomic principles or rules. Thus Peking man was named *Sinanthropus pekinensis*, a distinct genus and species, although this form is certainly closely related to the previously named Java man, *Pithecanthropus erectus*, and is at best not more than a separate species. Even fossil specimens of *Homo sapiens* have been given distinct species names, quite unjustifiably. Accordingly there exists in the literature a large number of such names, often several for one specimen, which do not relate to any formal classification and cannot be regarded as having the valid-

ity accorded such names in zoology or paleontology.

Line of development. General interpretations are hindered by incompleteness of data, which leaves room for considerable disagreement. The Australopithecinae survived to the end of the lower Pleistocene, in Asia as well as in Africa. The circumstances of advance to a euhominid status are quite unknown. It seems clear that the *Pithecanthropus* lineage (Java and Peking man) constituted the human population of the Far East in the later part of the lower and also the middle Pleistocene. This lineage may have extended to North Africa during the same time, judging from dental likeness to the fossil jaws found there. The lineage may also have survived in Java in the form of Solo man. The early Heidelberg mandible of Europe, however, presents dental and morphological differences (such as teeth which are distinctly more modern) which indicate a separate line of development.

Connections between the above and the populations of the upper Pleistocene (Solo man excepted) cannot be traced. Here, questions relate largely to connections between *Homo sapiens* and Neanderthal man (Rhodesian man being insufficiently known). By one school of opinion, the evolutionary separation between them is early, going back at least to the middle Pleistocene. This would mean that *Homo sapiens* as a general form has considerable antiquity. The view is based on interpretation of the Steinheim and Swanscombe skulls of the Second Interglacial, accepting the former as an early Neanderthal and the latter (assuming that the missing forehead was substantially of *sapiens* type) as early *Homo sapiens*. An intermediate view assumes that these two skulls were similar in nature and that they provide a middle Pleistocene source from which both Neanderthal man and *Homo sapiens* evolved in different directions. Both of these views recognize the Fontéchevade skulls (and perhaps the Kanjera skulls) as establishing the presence of *Homo sapiens* during the Third Interglacial at least.

A third view considers Steinheim and Swanscombe as basically Neanderthal, and holds that the still later Neanderthal population of the Third Interglacial was the common source from which the Fourth Glacial Neanderthals became somewhat more specialized and from which *Homo sapiens* arose by a late and fairly rapid progressive evolution. This view is based on the fact that the Third Interglacial and Fourth Glacial remains of Europe are almost exclusively Neanderthal in nature, and discounts the evidence of the Fontéchevade and Kanjera skulls. It also rests on the important Mount Carmel skeletons of Palestine, from the caves Et-Tabūn and Es-Skhūl. These are of Fourth Glacial date. A skeleton from the Tabūn cave is of basically Neanderthal type. Several from the Skhūl cave, believed to be slightly later in date, are of *sapiens* form except that they have large brow ridges and facial features recalling Neanderthals. In the above view this whole population represents a Fourth Glacial transformation of one segment of the Neanderthal stock into modern man, not long before its invasion of Europe, at the beginning of the upper Paleolithic during the Gottweig Warm Interstadial. Those opposed to this interpretation see *Homo sapiens* as earlier in origin and explain the Skhūl population as one resulting from mixture between already existing Neanderthal and *sapiens* men. [W.W.H.]

Bibliography: M. Boule and H. V. Vallois, *Fossil Men*, 1957; W. E. Le Gros Clark, *The Fossil Evidence for Human Evolution*, 1955; W. Howells, *Mankind in the Making*, 1959.

Fossil seeds and fruits

Seeds, ovules containing a fertilized egg and ready to be shed from the plant, are reproductive organs characteristic of both gymnospermous and angiospermous plants. In angiosperms an additional structure, the matured ovary, encloses one or more seeds to form a fruit.

Seeds and fruits are less commonly found as fossils than are vegetative remains. They may be preserved structurally as casts or as compressions which are sometimes found with leaf compressions. Seeds and fruits often occur in lignites.

The oldest known seed plants are of Mississippian age. Carboniferous seed plants include the extinct Cordaitales, probable conifer ancestors, and Pteridospermae, seed plants with fernlike foliage. Both groups had similar seeds basically like those of modern cycads. No embryos have been found in Paleozoic seeds. *See* CORDAITALES; PTERIDOSPERMAE.

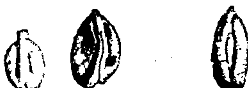
During the Mesozoic Era, all major modern groups of seed plants were represented, along with members of the declining cordaitalean and pteridospermous stocks. Among the most completely known Mesozoic seeds are those of the cycadeoids, extinct cycad relatives. Their silicified, beehive-shaped trunks may include shoots bearing numerous small seeds between thick scales. The Mesozoic Caytoniales, whose small seeds were borne in fleshy enclosures, were thought at one time to be ancestral angiosperms. Now they are recognized as gymnospermous forms, probably related to the pteridosperms.

Angiosperm fruits are rare in Lower Cretaceous beds; Upper Cretaceous fruits are known from North Africa, Long Island, New York, and elsewhere. Tertiary fruits and seeds have been found in numbers in the United States in the Brandon lignite of Vermont (see illustration) and in the Clarno formation of central Oregon. Silicified cones of araucarian conifers are known from Patagonia. Best-known European Tertiary fruits and seeds are from the brown coals of Germany and the Eocene London Clay formation of England.

Important paleobotanical findings resulting from study of fossil seeds and fruits include the



endocarps of *Symplocos*



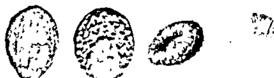
seeds of *Vitis* (grape)





seeds of fossil genus of family *Rutaceae*, related to *Phellodendron*





seeds of *Rubus* — endocarp of extinct species of *Nyssa* — endocarps of a species of *Alangium* — acorn cups, sp *Quercus*

Early Tertiary fruits and seeds found in Brandon lignite of Vermont. (*Elso S. Barghoorn*)

knowledge obtained of the independent evolution of the seed habit in unrelated groups, the discovery that much Carboniferous fernlike foliage was borne on seed plants rather than on ferns; and the discovery that *Glossopteris*, an important plant in widespread Permian floras of the Southern Hemisphere, was a seed plant. Pyritized fruits from the London Clay formation reveal the presence of many extinct genera along with modern genera in early Tertiary time. Morphological changes in herbaceous angiosperm seeds from sequences of Tertiary beds furnish data on rates of evolution in plants. Because plant classification is based primarily upon reproductive structures, fossil seeds and fruits provide highly reliable evidence for identification and interpretation of fossil plants. *See* PALEOBOTANY. [R.A.S.]

Bibliography: C. A. Arnold, *An Introduction to Paleobotany*, 1947; F. Kirchheimer, *Die Laubgewachse der Braunkohlenzeit*, 1957; E. M. Reid and M. E. J. Chandler, *The London Clay Flora*, 1933; R. A. Scott, Fossil fruits and seeds from the Eocene Clarno Formation of Oregon, *Palaeontographica*, 96(B):66–97, 1954.

Foucault pendulum

A pendulum or swinging weight, supported by a long wire, by which J. B. L. Foucault demonstrated [illegible] vertical direction. The bob is set swinging along a meridian in pure translation (no lateral or circular motion). In the Northern Hemisphere, the plane of swing appears to turn clockwise; in the Southern Hemisphere it appears to turn counterclockwise, the rate being 15 degrees times the sine of the local latitude per sidereal hour (*see* DAY). Thus, at the Equator the plane of swing is carried around by Earth and the pendulum shows no apparent rotation; at either pole the plane of swing remains fixed in space while Earth completes one rotation each sidereal day. *See* INERTIAL GUIDANCE SYSTEM; PENDULUM; SCHULER PENDULUM. [F.H.R.]

Bibliography: O. Struve, B. Lynds, and H. Pillans, *Elementary Astronomy*, 1959.

Foundations

The term foundation may be applied to (1) the rock or soil to which the dead load of the structure and applied live loads are transmitted, (2) the substructure which serves to distribute the loads from the superstructure to the bearing material, or (3) a combination of both. Substructure denotes the lower structural portion of the edifice used to transmit the total weight and applied live loads and forces to the supporting rock or soil.

Foundations may be classified by the supporting material, as rock-bearing or soil-bearing. They may be classified by type of substructure, as spread, continuous, cantilever, floating or raft, pile, caisson, or grade beam. They may also be classified by use, as pier, abutment, retaining wall, culvert, or underpinning.

Soil and rock are as important as substructures in foundation design, because satisfactory performance of the structure depends upon adequate strength for each to resist the forces to which they are subjected. For a discussion of the action of soils in foundations *see* SOIL MECHANICS.

Footings. This term denotes slabs or blocks that spread loads from one or more columns on soil or

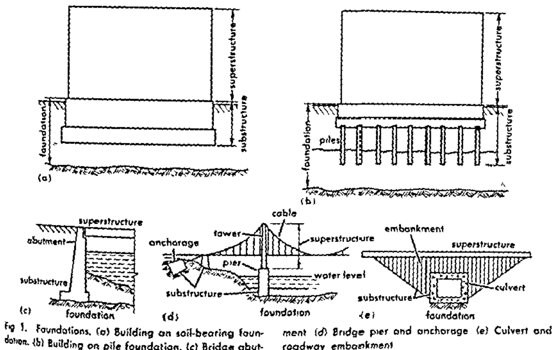


Fig 1. Foundations. (a) Building on soil-bearing foundation. (b) Building on pile foundation. (c) Bridge abutment (d) Bridge pier and anchorage (e) Culvert and roadway embankment

rock; it also includes pedestals. Footings may also denote continuous bearing slabs under walls.

Pedestal footings are solid rectangular concrete blocks (sometimes stepped), used where loads are light and small soil-bearing areas suffice. They are also used under heavy loads if the footing rests directly on bedrock.

Spread footings are isolated bearing slabs, in contrast to mats and continuous wall footings.

Combined footings are large concrete slabs under two or more columns instead of an isolated spread footing under each load. They are also used under pairs of columns subjected to uplift or to opposing or different horizontal stresses. They may be rectangular or trapezoidal to provide bearing areas required. Cantilever footings are combined footings in which one column and one end of the footing are placed against an adjacent building line or building; in effect, the end of the footing is cantilevered to carry the column.

Ribbed footings are combined footings stiffened by a longitudinal rib between pedestals.

Wall footings are continuous concrete walls resting on a wider base.

Grillage footings consist of two or more tiers of closely spaced structural steel beams resting on a concrete block, each tier being at right angles to the one below and made of steel encased in concrete.

Foundation walls. Foundation walls are continuous substructure walls or beams bearing on ground or rock, located under the superstructure of a building, and usually constructed of masonry or concrete. They may be shallow or deep, may bear continuously or span between footings, and may form a basement. Light structures without basements may be founded on simple walls, the depth depending on frost penetration and distance to adequate bearing material. A continuous footing may be needed to reduce unit bearing pressure.

Grade beams are concrete members supporting building walls and carrying their loads to individual footings located under columns or pilasters. This type of construction is adapted to buildings

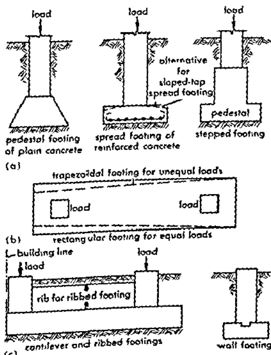


Fig. 2. Footings. (a) Isolated footings. (b) Combined footings. (c) Continuous footing.

having light superstructure walls where the supporting soil stratum is located at a depth of several feet, and where column loads are relatively heavy.

Continuous foundation walls may be used to support a series of concentrated loads. They may be desirable where the frost line would require deep grade beams, where soil-bearing values are locally variable, or where a high base course results in a deep member.

Basement walls are the walls enclosing an excavated space beneath the first floor of a building. They resist the lateral pressures of soil and water by spanning vertically between basement and first floor slabs or as cantilevers. Concrete walls are stiff distributing members for superimposed loads.

Waterproofing basements. The passage of exterior water through walls, floor, or construction joints may be prevented by the reduction of exterior hydrostatic pressure, integral tightness, membranes, and surface coatings. When the prevention of dampness and leakage through basement walls is necessary, the basement should be designed so that the water can be stopped at the outside surfaces. Coatings applied to the inner face are of doubtful efficacy. Dampness in basements due to condensation (sweating) will not be stopped by waterproofing, but usually will be removed by slight heating. Ventilation may remove dampness except when warm, moist air is brought into contact with cold walls.

Drainage lines placed outside the walls may reduce or remove water pressure, if they can discharge to a free-draining outlet. They should be of vitrified clay, cast iron, or other material that will not deteriorate. Pipes should have open joints or perforations that will let in water. Silting should be prevented by use of uncalked bell and spigot joints, tar paper wrappings, setting pipes in gravel, or similar means.

Integral tightness as a method of waterproofing requires use of as few joints as possible and special provisions at unavoidable joints. Dense concrete is required to prevent seepage. Admixtures may be beneficial in reducing porosity but will not stop leaks at cracks and joints.

Membrane waterproofing consists of layers of fabric and tar or asphalt placed to form a continuous wall coating, or wall and slab coating, with these blankets protected by coverings of masonry or concrete. A membrane should be composed of

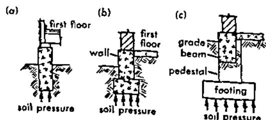


Fig. 3. Foundation walls. (a) Simple wall. (b) Wall and continuous footing. (c) Grade beam and footing.

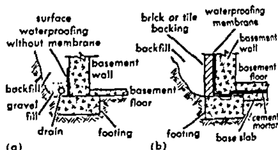


Fig. 4. Waterproofing basements. (a) Surface method with drainage line. (b) Membrane method.

at least two mopped, lapped plies of felt or fabric, and three to five plies if hydrostatic pressures are large.

Surface waterproofing consists of coatings of bituminous emulsion or mastic, without a membrane fabric or protective covering. Two successive moppings are desirable. Such coatings are beneficial, but cannot be relied upon to resist leakage permanently when under hydrostatic pressure for 1 day or longer.

Foundation mats. A continuous footing that supports an entire structure, such as a floor, is called a foundation mat, or raft foundation. These mats are generally used on soils of low bearing power where single footings would be so large that there would be little space between them, or where continuity is desired, as in basements subject to water pressure.

Uniform mats consist of a reinforced concrete slab of constant thickness, supporting walls, and columns. They are thick, rigid, and strong. They may be undesirable where applied loads are of unequal spacing and magnitude because such mats cannot spread loads far on compressible soil. Column spacings over 20 ft on centers may require an uneconomically thick mat. The unit shear value in the concrete due to column loads may require pedestals above the mat or depressed footings below.

Ribbed mats consist of a reinforced concrete slab stiffened by ribs above or below the mat, permitting the slab to be thinner than for a uniform mat. Ribs may run in one or both directions. The mat is designed as a system of one-way or two-way slabs supported against the ribs. Ribs may consist of depressed bands, concrete beams, or concrete basement walls. Use of rib beams above the mat requires lowering the mat to the underside of the beams, and providing compacted fill to carry the basement floor slab, in which piping may be run if sleeves are used through the ribs.

Floating foundations consist of mat and side walls set in an earth excavation. This type is used where major increase of pressure on the soil is undesirable. A weight of soil is excavated equal to the load from the structure, or such portion as is in excess of any allowable increase in pressure. This assumes that the soil has been subjected to its present pressures for so many years that practical

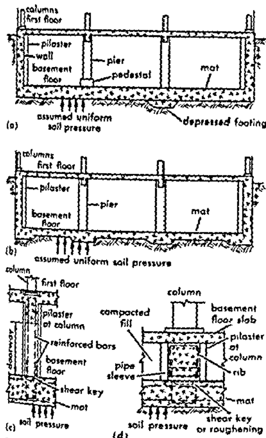


Fig 5. Foundation mats. (a) Uniform mat. (b) Ribbed mat with depressed band ribs. (c) Ribbed mat with cross walls (d) Ribbed mat with fill above mat

equilibrium has been reached and further consolidation will be too slow to be important. Waterproofing this basement is usually required.

Hydrostatic pressure acts on foundation mats wherever the ground-water level is higher than the bottom of the mat. It is resisted by the weight of the mat and the wall or column loads. The walls and mat are designed to resist lateral and upward pressures from the height of displaced water.

Underpinning. This is the addition of permanent supports to existing foundations. Underpinning is used to provide additional load-carrying capacity or to reduce settlements by enlarging or deepening the foundations. When excavation extends below foundations of adjacent structures, they usually require underpinning. Increase of column loads due to additional building stories or increased floor loads may necessitate underpinning. Building settlements during or after construction may require underpinning; these may be due to poor soil conditions not discovered by exploratory work, to faulty assumptions of soil properties and actions, or to changes occurring subsequent to construction, such as vibration, or changes in groundwater levels.

Underpinning operations are divided into two parts: (1) putting the structure and foundation into such condition that it is safe to install the underpinning, by the use of shores and needles, and (2) installation of the underpinning itself, consisting of enlarged footings, piles, piers, or caissons. Practical considerations are avoidance of loss of ground, which might cause settlements from soil movement; wedging and testing the underpinning to bring load onto it, to avoid settlement; and provision of bracing to prevent lateral movement.

Shoring is a method of supporting a member by inclined braces placed against walls or piers. Shores are made of timber or steel. They rest on foundation blocks, against which the load is taken up by wedging or by jacking. Shores are effective for light buildings, and occasionally for heavy ones. They are usually temporary. Permanent shores consisting of piles driven at a steep angle are sometimes used.

Needling is a method of supporting a member by transferring its load to temporary horizontal beams. Needles remove the load from walls or columns and are supported at the ends on suitable footings located away from the area to be underpinned. They are usually long, and are of timber, steel, or composites of steel and wood or steel and concrete. The load is transmitted by wedges or by screw jacks. Needles are most useful for light loads. *See* CAISSON FOUNDATION, PILE FOUNDATION; RETAINING WALL [R.D.C.]

Bibliography: P. Andersen, *Substructure Analysis and Design*, 2d ed., 1956; R. D. Chellis, *Pile Foundations*, 1951, C. W. Dunham, *Foundations of Structures*, 1950, W. C. Huntington, *Earth Pressures and Retaining Walls*, 1957, E. A. Prentis and L. White, *Underpinning, Its Practice and Application*, 2d ed., 1950; G. P. Tschebotarioff, *Soil Mechanics, Foundations, and Earth Structures*, 1951.

Four-bar linkage

A mechanism consisting of four elements, or links, which are constrained to move relative to each other in a determinable pattern, depending upon their configuration. The four bar linkage is the simplest complete mechanism that can consist entirely of lower, or closed, pairs. It is used more than any other mechanism, but it is capable of so many variations that it is frequently difficult to recognize. One link of the four-bar mechanism is fixed with respect to the ground or with respect to a machine frame.

Relative motion, or relative displacement, of one point in a linkage with respect to any other point in the linkage can be predicted throughout a cycle. A cycle of the mechanism is completed when a moving link has passed successively through all possible positions and has returned to its original position. The path of any point P with respect to any other point Q can be determined by fixing the link containing Q, then plotting successive positions of P as the linkage moves through a cycle. For relative displacements with respect to time (relative velocities), *see* VELOCITY ANALYSIS.

Equivalent linkage. Two linkages are equivalent when relative motions of points on given links are unchanged, although physical appearance of a corresponding link may be radically different between the two linkages. Links may be expanded, or extended in any direction at will, without changing the relative displacement of points in the linkage. The four-bar linkage of Fig. 1*a*, for example, is the kinematic equivalent of the linkage of Fig 1*d*, within limits of oscillation of link 2 shown by the dotted line. The transformation is made by expanding links 3 and 4, as in Fig. 1*b* The portion of expanded links 3 and 4 that are retained to produce the linkage of Fig 1*d* are indicated by shading in Fig. 1*c*. Recognition of an equivalent linkage will often simplify analysis of velocities and accelerations.

Inversion of linkage. A linkage is said to be inverted when the fixed link is freed and another link is fixed For example, the linkage of Fig. 2*a*, called a crank and rocker mechanism, becomes a drag link mechanism when link 1 is freed and link 2 is fixed, as in Fig. 2*b*. Relative motion of points within the mechanism is dependent only upon the basic linkage, no matter which link happens to be fixed. For a given four-bar linkage there are four possible inversions, with each of the four links fixed in succession.

Angular velocity. The rotation of a link about a point can be expressed as a scalar (not vector) quantity, symbol ω Customary units are: revolutions per minute, to express average angular velocity; degrees per second, for cam analysis; and radians per second, for conversion of instantaneous

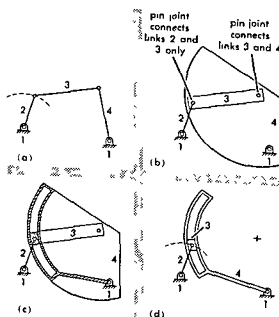


Fig. 1. Conversion of bar linkage (a) through intermediate forms (b) and (c) to equivalent slider linkage (d).

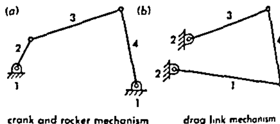


Fig 2. Changing fixed link from 1 (a) to 2 (b) inverts linkage.

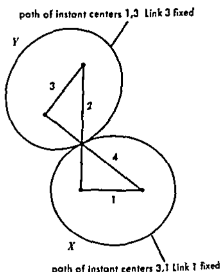


Fig 3. Four-bar linkage can be converted to rolling cams whose surfaces are the centrodes of opposite links.

angular velocity to tangential velocity. Angular velocity ratio of one link with respect to another in the same mechanism is frequently expressed as a fraction, as for example 2/1 or 3.7/1.

Dead center or dead point. In a crank and connecting-rod linkage, dead center is either of two positions in which the crank cannot be set in motion by a single force exerted at the opposite end of the connecting rod.

Centrode. Centrode is the path of instant centers of one link of a mechanism with respect to another. The centrodes of two moving links can be used as rolling curves that will provide an equivalent linkage, imparting to the two links unchanged relative motion. In Fig. 3, oval figure X is the centrode of link 3, constructed by fixing link 1 in the paper; figure Y is the centrode of link 1, constructed by fixing link 3 in the paper. Rolling of X on Y will give an equivalent linkage, links 2 and 4 being no longer necessary.

Crank and rocker mechanism. A four-bar linkage (Fig. 2*a*) consisting of crank, connecting rod, rocker, and fixed link is used for indexing devices, lubricators, and generally where the rocker is caused to oscillate periodically by a constant rotation of the crank. Endlessly adaptable, by changing the proportions of the links, the crank and rocker

linkage is also used in many devices in which the crank is not designed for a complete revolution.

Drag link mechanism. This four-bar linkage (Fig. 4*a*), is an inversion of the crank and rocker mechanism. The drag link coupling, for which the mechanism is named, was used by James Watt (1736–1819) to couple parallel but slightly offset shafts. Figure 4*b* shows the drag link *L*. The axes of shafts *R* and *B* are shown as pivot points *R* and *B* in Fig. 4*a*.

Parallel crank linkage. In a parallel crank linkage $\overline{AB} = \overline{CD}$ and $\overline{BC} = \overline{AD}$. The side rod on a stream locomotive is link 3 of Fig. 5*a*, links 2 and 4 being the wheels and link 1 the locomotive frame.

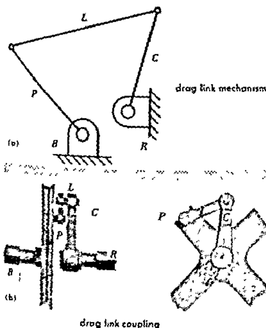


Fig. 4. Drag link for coupling parallel shafts. (a) Schematic. (b) Coupling. (G. Rennie, *Illustrations of Mill Work, 1841*)

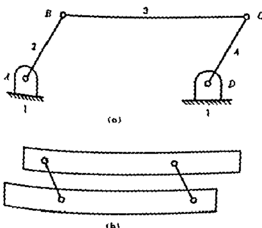


Fig. 5. Parallel crank linkage (a) is basis of parallel rule (b).

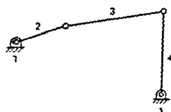


Fig. 6. Slow-motion linkage.

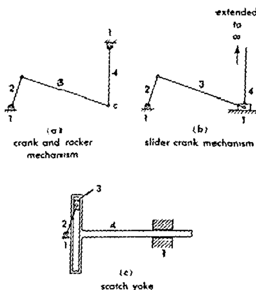


Fig. 7. (a) Crank and rocker mechanism. (b) Slider crank mechanism, with one bar of infinite length, is equivalent to scotch yoke (c).

A parallel ruler (Fig. 5*b*) employs this linkage (*see* DRAFTING MACHINE; STRAIGHT LINE MECHANISM).

An equal crank linkage is similar to the parallel crank linkage, except connecting rod 3 is shorter than fixed link 1. A common application of the linkage is to the conventional automobile steering mechanism (*see* ACKERMAN STEERING). This linkage was patented as an improvement in axletrees by Rudolph Ackermann in 1818.

Slow motion linkage. This crank and rocker mechanism is so named because of the relatively small angular displacement of link 4 for a large angular displacement of link 2 at or near the position shown in Fig. 6. A variation of this linkage was used in the Corliss steam engine valve gear (by G. H. Corliss in 1849) to provide quick opening and closing with slow motion near the end of valve oscillation. The rotary valve was pinned to link 4.

Sliding pair and linkage. If the length of the rocker of the crank and rocker mechanism (Fig. 7*a*) is increased to infinity, the path of point *C*, at the end of connecting rod 3, becomes a straight line (Fig. 7*b*). Rocker 4 can then be replaced by a sliding block, or crosshead, which is confined to straight reciprocating motion by a slide, or guide. The slider crank mechanism, thus formed, is widely used. By inverting the slider crank mechanism ar

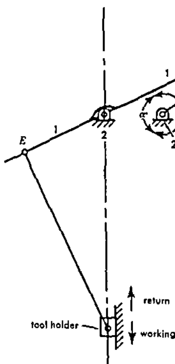
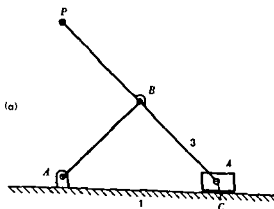


Fig. 8. Whitworth quick return mechanism.



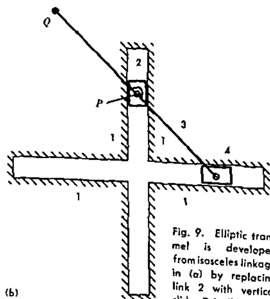
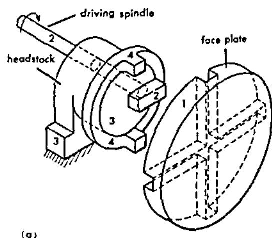


Fig. 9. Elliptic trammel is developed from isosceles linkage in (a) by replacing link 2 with vertical slider 2 in (b).

by making other alterations, a variety of useful mechanisms is obtained. Several of these are described below. *See* SLIDER CRANK MECHANISM.

Sliding slot linkage or scotch yoke. An alteration of the slider crank mechanism in which the length of the connecting rod (link 3, Fig. 7*b*) is effectively increased to infinity, is shown in Fig. 7*c*. Link 4 has simple harmonic motion when the angular velocity of crank 2 is constant.

Quick return mechanism. The usual quick return mechanism is an inversion of the slider crank mechanism in which the crank, link 2, is fixed (Fig. 8).



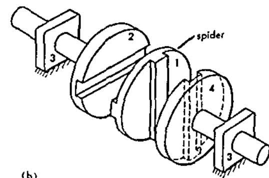


Fig 10. Oldham's coupling (a) and elliptic chuck (b).

Joseph Whitworth (1803–1887) specified the linkage in a vertical shaper patent in 1849. The tool holder completes its working stroke in θ degrees and its return stroke in α degrees of rotation of driving link 3. If link 3 has constant angular velocity, the return stroke will take about half as much time as the working stroke. The length of stroke of the tool holder block can be altered by moving pivot E along link 1.

Isosceles linkage. If the crank and connecting rod of the slider crank mechanism are made equal in length, the mechanism of Fig 9*a* results. Triangle ABC is at every position an isosceles triangle. Point P, on the extension of connecting rod 3, traverses a vertical path. If P is guided in a vertical line (Fig 9*b*), crank 2 can be eliminated. The re-

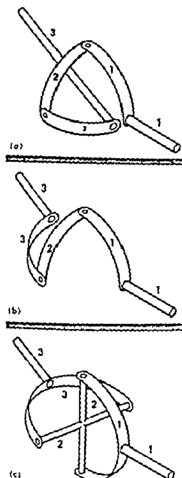


Fig 11. Conic drag link coupling (a) can be modified (b) into Hooke's joint (c).

sulting mechanism is the elliptical trammel, so named because point Q on an extension of link 3 traces an elliptical path.

Oldham's coupling (John Oldham, fl. 1810), for coupling parallel but slightly offset shafts, is obtained by fixing link 3 of Fig. 9b. Links 2 and 4 turn with the shafts being coupled, while link 1, carrying the crossed slots, becomes the spider (Fig. 10a). Angular velocities of the two shafts are equal, but offset of the center lines must be small. The elliptic chuck (Fig. 10b), for turning ovals in a lathe, is also obtained by fixing link 3 of Fig. 9b. Link 2 turns with the lathe spindle, and link 1, carrying the crossed slots, becomes the face plate upon which the chuck, carrying the work, is mounted. The device is made practical by expanding link 4 to surround link 2.

Conic linkage. In this mechanism, links are laid out on the surface of a sphere, with the axes of all joints converging at the center of the sphere (see UNIVERSAL JOINT). If the drag link mechanism is altered to obtain a conic linkage (Fig. 11a), it is only necessary to expand links 2 and 3 to obtain Hooke's joint (Fig. 11b and 11c). This is the basic mechanism of the universal joint. Robert Hooke (1635–1703) and Girolamo Cardano (1501–1576), are credited with describing the universal joint, although it is likely that Cardano recognized the linkage only as gimbals. The joint is known upon the continent of Europe as a cardan joint. See MECHANISMS. [E.S.F.]

Fourier series and integrals

Mathematical tools for the study of periodic phenomena, indispensable in the theory of wave motion, for example light waves and sound waves, oscillatory mechanical systems, such as vibrating strings, and celestial orbits. Fourier series and the related topics discussed below have important applications to other branches of mathematics of which the theories of probability and partial differential equations are worthy of special mention. Finally, the disciplines motivated by the subject itself enjoy a prominent position in pure mathematical research.

A function f of a real variable is said to be periodic with period T if $f(t+T) = f(t)$ for each t. The simplest examples of functions with a given period T are pure harmonics, that is functions of the form $f(t) = a_n \cos n\omega t + b_n \sin n\omega t$ where $\omega = 2\pi T^{-1}$ is the fundamental frequency and a_n, b_n are constants. The basic idea for the applications of Fourier series is that any function f of period T which satisfies rather mild restrictions can be represented as a superposition of pure harmonics:

$$f(t) \sim \sum_{n=0}^{\infty} (a_n \cos n\omega t + b_n \sin n\omega t)$$

or, in a more convenient form, using complex exponentials,

$$f(t) \sim \sum_{-\infty}^{\infty} c_n e^{in\omega t} \qquad (1)$$

If term-by-term integration of (1) is assumed to be legitimate, an easy calculation shows that

$$c_n = T^{-1} \int_{-T/2}^{T/2} f(t) e^{-in\omega t}\, dt \qquad (2)$$

(The interval of integration could be any interval of length T.) This motivates the formal definition of Fourier series. Suppose that f is a function of period T such that

$$\int_{-T/2}^{T/2} |f(t)|\, dt$$

exists and is finite. The coefficients $\{c_n\}$ defined by Eq. (2) are called the Fourier coefficients of f, and the series in formula (1), the Fourier series expansion of f. The coefficients uniquely determine the function; that is, if $c_n = 0$ for each n, then f is essentially the zero-function. In addition, many formal operations in which the series is substituted term-by-term for the function can be justified. Two important questions present themselves immediately.

Let

$$s_N(t) = \sum_{-N}^{N} c_n e^{in\omega t}$$

be the Nth partial sum of the Fourier series of f. The first question is, does $s_N(t)$ converge to $f(t)$ as N approaches ∞? The second question is, given a sequence $\{c_n\}$, is it the sequence of Fourier coefficients of some function?

The Fourier series of a continuous function need not converge everywhere. If t_0 is a given point, the convergence of $s_N(t_0)$ to $f(t_0)$ depends on the behavior of $f(t)$ for t in the neighborhood of t_0. However, if one takes the averaged partial sums,

$$\sigma_N = (N+1)^{-1}\sum_{k=0}^{N} s_k$$

then for continuous f, $\sigma_N \to f$ uniformly. The knowledge alone of the ordinary convergence of a Fourier series is of little importance in applications because, for purposes of calculation, it is necessary to know something about the rapidity of convergence. An example of a theorem treating this is the following: suppose $|df/dt| \leq M$ everywhere; then $|f(t) - s_N(t)| \leq \frac{1}{2}\pi M(N+1)^{-1}$.

The Riemann-Lebesgue lemma asserts that if $\{c_n\}$ is the sequence of Fourier coefficients of an integrable function then $c_n \to 0$ as $n \to \pm\infty$. The converse is false: not all trigonometric series

$$\sum_{-\infty}^{\infty} c_n e^{in\omega t}$$

with coefficients tending to zero are Fourier series.

To obtain a really satisfactory theory, one modifies the above questions and restricts attention to a special class of functions, namely the class L^2 *of* those functions f of period T for which the integral

$$\int_{-T/2}^{T/2} |f(t)|^2\,dt$$

exists and is finite. L^2 is a Hilbert space with the inner product of two functions f, g being

$$(f,g) = T^{-1}\int_{-T/2}^{T/2} f(t)\bar{g}(t)\,dt$$

(the bar denotes complex conjugation). In lieu of ordinary convergence, convergence in the mean exists; that is,

$$\lim_{N\to\infty}\int_{-T/2}^{T/2} |f(t) - s_N(t)|^2\,dt = 0$$

The partial sums s_N can be characterized by an extremal property: among all functions p of the form

$$p(t) = \sum_{-N}^{N} a_n e^{in\omega t}$$

the minimum of the quadratic deviation

$$\int_{-T/2}^{T/2} |f(t) - p(t)|^2\,dt$$

is achieved uniquely by $p = s_N$, the Nth partial sum of the Fourier series of f. For functions of the class L^2, there is the Parseval identity

$$T^{-1}\int_{-T/2}^{T/2} |f(t)|^2\,dt = \sum_{-\infty}^{\infty} |c_n|^2 \tag{3}$$

If f and g are two functions of L^2 with Fourier coefficients $\{c_n\}$ and $\{d_n\}$ respectively, then

$$T^{-1}\int_{-T/2}^{T/2} f(t)\bar{g}(t)\,dt = \sum_{-\infty}^{\infty} c_n\bar{d}_n \tag{3'}$$

Let l^2 denote the ensemble of sequences of complex numbers $\{c_n\}$ such that

$$\sum_{-\infty}^{\infty} |c_n|^2$$

is finite. One of the implications of Eq. (3) is that the Fourier coefficients of a function of class L^2 belong to l^2. The truth of the converse is the Riesz-Fischer theorem: If $\{c_n\}$ is a sequence of class l^2, then the trigonometric series

$$\sum_{-\infty}^{\infty} c_n e^{in\omega t}$$

is the Fourier series of a function of period T belonging to the class L^2. The statements of this paragraph are not peculiar to Fourier series, but rather they describe the general situation in a series expansion in terms of orthogonal functions. The set of functions $\{e^{in\omega t}\}$, in which n ranges over all integers, forms a complete orthonormal set in the L^2-space of functions on an interval of length T.

Almost periodic functions. These are a generalization of periodic functions. Suppose f is periodic with period T_0; then $f(t + nT_0) = f(t)$ for each integer n. In particular, in any interval of length greater than T_0, there is at least one number T of the form $T = nT_0$ so that $f(t + T) = f(t)$. A continuous function f is said to be uniformly almost periodic if, given a positive number ε, there is another positive number T, such that in any interval of length greater than T, there is at least one number T for which $|f(t + T) - f(t)| < \varepsilon$, for all t. An almost periodic function f has a almost periodic Fourier series

$$f(t) \sim \sum_{\nu} c_\nu e^{i\omega_\nu t} \tag{1a}$$

If f is actually periodic with fundamental frequency $\omega_0 = 2\pi T_0^{-1}$, then each ω_ν for which $c_\nu \neq 0$ will be an integral multiple of ω_0; in the general case, the series in formula (1*a*) is to be regarded as a superposition of pure harmonics without a common period. The generalization of Eq. (2) giving the formula for the almost periodic Fourier coefficients is

$$c_\nu = \lim_{T\to\infty} T^{-1}\int_{-T/2}^{T/2} f(t)e^{-i\omega_\nu t}\,dt \tag{2a}$$

In the above formula, the interval of integration can be taken to be any interval of length T, the limit always exists, and the limit is different from zero only for a countable number of frequencies ω_ν. If f is uniformly almost periodic, suitable combinations of the partial sums in formula (1*a*) converge uniformly on the whole real line to f. Conversely, a uniform limit of trigonometric poly

nomials, that is, finite sums of the form given in formula (1a), is a uniformly almost periodic function. There are also classes of discontinuous almost periodic functions, for example, the generalization of L^2 to the almost periodic case, which will not be discussed here. However, formula (3) has a simple analog:

$$\lim_{T\to\infty} T^{-1} \int_{-T/2}^{T/2} |f(t)|^2\, dt = \Sigma |c_r|^2 \qquad (3a)$$

Trigonometric integrals. The class of functions which admit a representation in the form (1a) is still too narrow for many purposes. It is necessary to admit continuous sums, or integrals, as well. A generalization is given by using Stieltjes integrals

$$f(t) \sim \int_{-\infty}^{\infty} e^{i\omega t}\, d\sigma(\omega) \qquad (4)$$

The meaning of formula (4) will be left rather vague; some examples follow.

If σ is of bounded variation on the whole real line, the integral above exists in the ordinary sense and σ is uniquely determined by its Fourier-Stieltjes transform f. A case of particular importance in probability is when σ is a distribution function, that is, σ is an increasing function with $\sigma(-\infty) = 0$ and $\sigma(+\infty) = 1$. If X is a real-valued random variable, then $\sigma(\omega)$ = probability $(X < \omega)$ is a distribution function. In this situation, f is called the characteristic function of the probability distribution. An important property of characteristic functions is that they are positive definite by which is meant that for each finite collection $(c_1,t_1), \ldots, (c_n,t_n)$ of pairs of complex numbers c_k and points t_k one has

$$\sum_{j,k=1}^{n} c_j\bar{c}_k f(t_j - t_k) \geq 0$$

Bochner's theorem states: A function f of a real variable is a characteristic function, that is, it is representable in the form (4) with σ a distribution function, if and only if (i) f is continuous, (ii) $f(0) = 1$, and (iii) f is positive definite. Characteristic functions are very useful in probability; the techniques involved are valid whenever one has to do with Fourier-Stieltjes transforms. These techniques are described below.

If X_1 and X_2 are two random variables with distribution functions σ_1 and σ_2 respectively, the random variable $X_3 = X_1 + X_2$ has the distribution function σ_3 defined by

$$\sigma_3(\omega) = \int_{-\infty}^{\infty} \sigma_2(\omega - \xi)\, d\sigma_1(\xi)$$

the convolution of σ_1 and σ_2. More generally, if σ_1 and σ_2 are two functions of bounded variation, their convolution σ_3, as defined above, is also of bounded variation. The Fourier-Stieltjes transform of a convolution is the product of the Fourier-Stieltjes transforms, that is, $f_3 = f_1 f_2$. Thus, the rather complicated operation of convolution can be replaced by the simpler operation of multiplication. Another useful principle concerns what are called limit theorems. Suppose $\{X_n\}$ is an infinite sequence of random variables with X_n having the distribution function σ_n and the characteristic function f_n. The question arises as to whether there is a limiting distribution. Not only are the calculations generally simplified by using characteristic functions, but also rigorous criteria are available; for example, if the sequence $\{f_n\}$ converges uniformly in some interval containing 0, then

$$\lim_{n\to\infty} f_n(t) = f(t)$$

exists for every t, f is also a characteristic function, corresponding to a distribution function σ, and

$$\lim_{n\to\infty} \sigma_n(\omega) = \sigma(\omega)$$

at each point of continuity of σ.

Proceeding purely formally from the representation in formula (4), differentiating n times gives

$$\frac{d^n}{dt^n} f(t) \sim i^n \int e^{i\omega t}\omega^n\, d\sigma(\omega)$$

Assuming that the appropriate integrals are absolutely convergent, one can calculate the moments of the function of bounded variation σ from the formula

$$\int \omega^n\, d\sigma(\omega) = i^{-n} \frac{d^n}{dt^n} f(0)$$

However, a more significant application of the device is that if L is a differential operator with constant coefficients, then at least heuristically,

$$Lf(t) \sim \int e^{i\omega t}\lambda(\omega)\, d\sigma(\omega)$$

where λ is the polynomial determined by

$$L_t e^{i\omega t} = \lambda(\omega) e^{i\omega t}$$

More generally, if L is any linear operation which commutes with translations, that is $LU(\tau) = U(\tau)L$ for each τ where $U(\tau)$ is defined by $(U(\tau)f)(t) = f(t+\tau)$, then $L_t e^{i\omega t} = \lambda(\omega) e^{i\omega t}$, and the situation is analogous. The idea is that the solution f of the equation $Lf = g$ where $g(t) \sim \int e^{i\omega t}\, d\rho(\omega)$ should be representable in the form (4) where $\lambda(\omega)\, d\sigma(\omega) = d\rho(\omega)$, and thus the problem is formally solved by division: $d\sigma(\omega) = d\rho(\omega)/\lambda(\omega)$. However, to get a useful theory, one has to admit generalized functions σ which are not of bounded variation and may not even be functions in the ordinary sense. A rigorous discussion of the representation of functions by generalized Fourier transforms is not only beyond the scope of this article but also a subject which presents challenging problems in mathematical research.

Fourier integrals. There is one class of functions for which a completely satisfactory theory of Fourier transforms exists. That is the generalization of the L^2-theory of Fourier series to Fourier integrals. It is customary to use L^2 to denote the Hilbert space of functions f such that

$$\int_{-\infty}^{\infty} |f(t)|^2 \, dt$$

exists and is finite with the inner product

$$(f,g) = (2\pi)^{-1} \int_{-\infty}^{\infty} f(t)\bar{g}(t) \, dt$$

For a function $f(t)$ in L^2, it is possible to form a Fourier transform $F(\omega)$ which is again a function in L^2. The function f will then have the representation (4) with $d\sigma(\omega) = F(\omega)\, d\omega$. That is

$$f(t) \sim \int_{-\infty}^{\infty} e^{i\omega t} F(\omega) \, d\omega \qquad (1b)$$

F is computed from the formula

$$F(\omega) = \lim_{T\to\infty} (2\pi)^{-1} \int_{-T}^{T} f(t) e^{i\omega t} \, dt \qquad (2b)$$

where the limit need not converge in the ordinary sense, but only in the mean. Otherwise, the theorems about Fourier series usually have appropriate generalizations. For example, if one forms the partial sums

$$s_N(t) = \int_{-N}^{N} e^{i\omega t} F(\omega) \, d\omega$$

one again has convergence in the mean:

$$\lim_{N\to\infty} \int_{-\infty}^{\infty} |f(t) - s_N(t)|^2 \, dt = 0$$

The Parseval identity takes the form

$$(2\pi)^{-1} \int_{-\infty}^{\infty} |f(t)|^2 \, dt = \int_{-\infty}^{\infty} |F(\omega)|^2 \, d\omega \qquad (3b)$$

If f and g are two functions of L^2 with Fourier transforms F and G respectively, then

$$(2\pi)^{-1} \int_{-\infty}^{\infty} f(t)\bar{g}(t) \, dt = \int_{-\infty}^{\infty} F(\omega)\bar{G}(\omega) \, d\omega \qquad (3'b)$$

and in particular, for the convolution

$$(2\pi)^{-1} \int_{-\infty}^{\infty} f(t-s)g(s) \, ds = \int_{-\infty}^{\infty} e^{i\omega t} F(\omega) G(\omega) \, d\omega \qquad (5)$$

The preceding discussion has been confined to functions of one variable, but the theory is easily extended to functions of several variables. The results persist except for trivial modifications, if one considers t and ω as n-dimensional vectors, $t = (t^{(1)}, \ldots, t^{(n)})$, $\omega = (\omega^{(1)}, \ldots, \omega^{(n)})$, and ωt is replaced in the exponentials by the inner product $\omega^{(1)}t^{(1)} + \cdots + \omega^{(n)}t^{(n)}$. Here one is dealing with multiple Fourier series and multiple Fourier transforms. The discussion of representations of the form (4) for functions of several variables is of particular interest in the study of partial differential equations with constant coefficients. See ORTHOGONAL POLYNOMIALS; PROBABILITY; SERIES. [C.S.H.]

Bibliography: S. Bochner, *Lectures on Fourier Integrals*, Annals of Math. Studies, 42, 1959; E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, 2d ed., 1950; A. Zygmund, *Trigonometrical Series*, 1955.

Fourth dimension

The designation frequently given to time in the theory of relativity. This terminology is intended to convey the idea that, in relativity, time and its passage are not absolute categories unrelated to the state of motion of the observer and hence common to all, but that time forms but part of a whole, usually referred to as the space-time continuum.

Just as in ordinary space it is meaningful to speak of the value of the x component of a given line segment only if the coordinate system is identified with respect to which the x component is to be determined, so the time interval between two specified events is unequivocally determined only if the state of motion of the observer performing the measurements is given.

For a more detailed discussion see RELATIVITY; SPACE-TIME. [P.G.B.]

Fox

Any of a number of small to moderately large, doglike, carnivorous mammals of the family Canidae, found throughout much of North Africa, Eurasia, and North America. Foxes are alike in having long, bushy tails, and the pupils of the eyes elliptical (they are round in wolves and coyotes). Most foxes have short legs and long, pointed ears.



The red fox, Vulpes fulva; length about 25 in (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

Most widely distributed and best known is the red fox, *Vulpes fulva*, common over most of North America and in Eurasia. It is valued both for its fur and for the sport of the chase. The somewhat larger gray fox, *Urocyon cineroargenteus*, is an American forest fox; the *Vulpes* foxes tend to prefer open country. *Alopex lagopus*, the Arctic or blue fox, is a valuable fur animal of the northland. See CARNIVORA. [J.D.B.]

Fractionating column

An apparatus used widely for countercurrent contacting of vapor and liquid to effect separations by distillation or absorption. In general, the apparatus consists of a cylindrical vessel with internals designed to obtain multiple contacting of ascending vapor and descending liquid together with means for introducing or generating liquid at the top and vapor at the bottom.

Figure 1 shows such a column that can be applied to distillation. A vapor condenser is used to produce liquid (reflux) which is returned to the top, and a liquid heater (reboiler) is used to generate vapor for introduction at the bottom. In a simple absorber, the absorption oil is the top liquid and the feed gas is the bottom vapor. In all cases, changes in composition produce heat effects and volume changes so that there is a temperature gradient and a variation in vapor and liquid flows from top to bottom of the column. These changes affect the internal flow rates from point to point throughout the column and must be taken into account in designing the column.

Fractionating columns used in industrial plants range *in diameter* from a few inches to 40 ft and in height from 10 to 200 ft. They operate at pressures as low as a few millimeters of mercury and as high as 3000 psi; at temperatures from −300 to 700°F. They are made of steel and other metals, of ceramics and glass, and even of such materials as bonded carbon and plastics.

A variety of internal devices has been used to obtain more efficient contacting of vapor and liquid

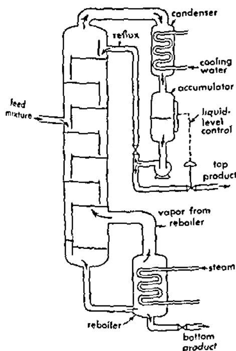
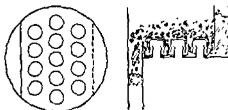


Fig 1. Elements of a *fractionating column*.



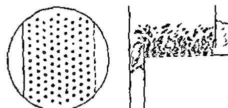






Fig 2. Contacting devices for fractionating column

The most widely used devices are the bubble-cap plate, the perforated or sieve plate, and the packed column (Fig. 2).

Plates. The bubble-cap plate is a horizontal deck with a large number of chimneys over which circular or rectangular caps are mounted to channel and distribute the vapor through the liquid. Liquid flows by gravity downward from plate to plate through separate passages known as downcomers. The number of caps and the size of downcomers are designed to fit the expected internal flow rates of liquid and vapor.

The bubble-cap plate is the most common type of contacting device used in fractionating columns. It can handle a wide range of vapor and liquid flows, is not obstructed by small amounts of rust and other solids, and is manufactured by a large number of suppliers.

The perforated or sieve plate is a horizontal deck with a multiplicity of round holes or rectangular slots for distribution of vapor through the liquid. The sieve plate can be designed with downcomers similar to those used for bubble-cap trays, or it can [illegible] so that both liquid [illegible] restricted operating range so that [illegible] at

vapor flows much less than the design rate results in poor contacting.

Recently a number of variations and improvements have been made in the sieve tray to improve the operating range. The most successful of these is the valve tray in which a check valve is placed over each perforation to adjust for the variation in vapor flow. The valve usually consists of a disk over each perforation which is held within a cage attached to the deck The disk is free to move up and down within the limits imposed by the cage and thus adjusts to the vapor flow

Although it is older than the bubble-cap plate, the sieve tray did not obtain wide use until stainless-steel sheets became readily available to overcome difficulties with corrosion and fouling. With the newer improvements, such as the valve tray, this type of plate promises to supplant the bubble-cap plate in many services

Packings. The packed column is a bed or succession of beds made up of small solid shapes over which liquid and vapor flow in tortuous countercurrent paths. Some of the more common types of packing are shown in Fig. 2 Expanded metal or woven mats are also used as packing The packed column is used without downcomers, but in larger sizes usually has horizontal redistribution decks to collect and redistribute the liquid over the bed at successive intervals of height. The packed column is widely used in laboratories It is often used in small industrial plants, especially where corrosion is severe and ceramic or glass materials must be chosen.

Other contacting means sometimes used in fractionating columns are the disk-and-doughnut tray for very viscous or dirty liquids, the spiral ribbon for very small columns, the spray column, and the welled-wall column. *See* DISTILLATION; GAS ABSORPTION OPERATIONS. [M.SO.]

Bibliography: C. H. Nielson, *Distillation in Practice*, 1956; J. H. Perry (ed.), *Chemical Engineers' Handbook*, 3d ed., 1950.

Frame of reference

A base to which to refer physical events. A physical event occurs at a point in space and at an instant of time. Each *reference* frame must have an observer to record events, as well as a coordinate system for the purpose of assigning locations to each event. The latter is usually a three-dimensional space coordinate system and a set of standardized clocks to give the *local time* of each event. For a discussion of the geometrical properties of space-time coordinate systems, *see* SPACE-TIME; *see also* RELATIVITY.

Newtonian reference frames. In the ordinary range of experience, where light signals, for all practical purposes, propagate instantaneously, the time of an event is quite distinct from its space coordinates, since a single clock suffices for all observers, regardless of their state of relative motion. The set of reference frames which have a common clock or time is called Newtonian, since Isaac Newton regarded time as having invariable significance for all observers.

Inertial reference frames. In these a body moves with constant velocity when free of impressed forces. It follows that one inertial frame cannot be accelerating with respect to another. If x_1, y_1, z_1 and x_2, y_2, z_2 are the coordinates of a particle P in parallel inertial frames 1 and 2, where 2 is moving with respect to 1 with velocity components v_x, v_y, v_z, then

$$x_2 = x_1 - v_x t \qquad y_2 = y_1 - v_y t \qquad z_2 = z_1 - v_z t$$

(Newtonian frames are understood here so that $t_2 = t_1$.) This is called a Galilean transformation. Since the vs are independent of time, the component accelerations of P in 1 and 2 are the same, that is,

$$A_{x2} = d^2x_2/dt^2 = d^2x_1/dt^2 = A_{x1} \quad \text{etc.}$$

Thus, Newton's second law, $F = mA$, where m is the mass of the body, holds in all inertial frames. *See* GALILEAN TRANSFORMATIONS.

Accelerated reference frames. A noninertial frame of reference is called an accelerated frame. A point fixed in such a frame will be accelerated with respect to inertial frames and so must be held by a force. An observer in an accelerated frame will see accelerations that are not proportional to the impressed forces, since a force is required to keep the particle at rest. Newton's second law does not hold without modification.

Rotating reference frames. A rotating frame of reference is one whose coordinate axes are rotating with respect to some inertial system. A point P fixed in the rotating frame moves uniformly about a circle in the inertial frame and is accelerated towards the axis of rotation. (For further details, *see* CENTRIPETAL FORCE.) There is an additional acceleration (with respect to the inertial system) if P moves in the rotating system. For example, if the point moves away from the axis along a moving radius, it gains circular speed also. For additional details, *see* CORIOLIS ACCELERATION AND FORCE.

Newton's second law can be modified to take account of these two accelerations that are not observed in the rotating system. Let A_{Ae}, A_C, A_{Co} be the observed, centripetal, and Coriolis accelerations, respectively. The total acceleration in the inertial frame is

$$A = A_{Ae} + A_C + A_{Co}$$

where the accelerations must be added vectorially. Defining the centrifugal force, $F_C = -mA_C$, and the Coriolis force, $F_{Co} = -mA_{Co}$, enables Newton's second law to be written

$$mA_{Ae} = F + F_C + F_{Co}$$

In rotating systems, a consistent mechanical theory is obtained if the two additional forces are added to the impressed forces Conversely, the presence of

such forces will enable the observer to determine the motion of his reference frame.

Astronomical reference frames. These are fixed relative to celestial objects. The most convenient one for observational purposes is, of course, a geocentric system. The earth's axis of rotation is used as a coordinate axis. A second reference direction is chosen as the intersection of the earth's equatorial plane and the plane of the earth's orbit (the ecliptic). The basic time unit is the earth's rotation period, the sidereal day. The positions and times of distant astronomical events are determined indirectly. *See* ASTRONOMICAL COORDINATE SYSTEMS.

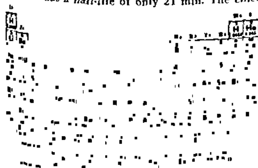
Reference frames are regarded as more basic as more phenomena appear regular with respect to them. Thus the planetary motions are described naturally in an ecliptic coordinate system centered on the sun, and the motions of stars as viewed from the earth appear regular in a coordinate system with its origin at the center of our galaxy.

Reference frames based on the mean positions of large enough celestial objects, for example, a group of galaxies, approximate inertial frames closely. Between two such reference frames, the distances and relative velocities may be so large that transmission of light signals takes considerable time. Newtonian reference frames with a single universal time are no longer meaningful, and each reference frame must have its own system of time. Inertial frames of this type are called Galilean reference frames, a somewhat misleading name since the transformation of the coordinates of an event from one such frame to another is not the Galilean transformation given earlier (with $t_1 = t_2$) but one which involves the time directly. It is known as a Lorentz transformation. Galilean reference frames are often called Lorentz frames. *See* LORENTZ TRANSFORMATIONS. [B.C.]

Bibliography: A. Einstein, *The Meaning of Relativity*, 5th ed., 1955; W. S. Krogdahl, *The Astronomical Universe*, 1952.

Francium

Element 87, symbol Fr, an alkali element filling the position below cesium in the first group in the periodic table of elements. Francium is distinguished chiefly by its nuclear instability, since it exists only in short lived radioactive forms the most durable of which has a half-life of only 21 min. The chief



isotope of francium is actinium-K, an isotope of mass 223, which arises from the radioactive decay of the element actinium. Actinium in turn is a decay product of the rare uranium isotope, U^{235} and hence is a normal component of uranium minerals. Actinium decays chiefly by β-decay, but the French scientist, Marguerite Perey, discovered that 1% of the disintegrating actinium atoms emit alpha particles and are thus converted to atoms of element 87 of mass number 223. Actinium-K emits beta particles with a half-life of 21 min. Perey's experiments with this isotope constituted the first reliable identification of an isotope of element 87; she was thus credited with the discovery of the element for which she chose the name francium.

Other isotopes of francium, none as long-lived as AcK, have been synthesized by the bombardment of thorium targets with high-energy protons, deuterons, or helium ions. From the properties of the known isotopes and from the systematic trends in nuclear stability, it is reasonably certain that no long-lived form of element 87 will ever be found in nature or synthesized artificially.

The chemical properties of francium can be studied only on the tracer scale. The element has all the properties expected of the heaviest alkali element. With few exceptions, all the salts of francium are water soluble. Francium remains in aqueous solution when other elements are precipitated as insoluble hydroxides, fluorides, sulfides, or chromates. Radiochemical purification of francium can be achieved simply, by removing actinium and other foreign radioactivity by successive precipitation of a series of insoluble scavenger precipitates, following which the radiations of francium can be sought in the supernatant solution. Another useful method involves the coprecipitation of the tracer amounts of francium with added cesium carrier which is precipitated from solution as the insoluble perchlorate, chloroplatinate, or silicotungstate. Francium can also be separated by use of cation-exchange resins. *See* ALKALI METALS [E.K.HY.]

Bibliography: K. W. Bagnall, *Chemistry of the Rare Radioelements*, 1957.

Franck-Condon principle

In any molecular system the transition from one energy state to another occurs so near to instantaneously that the nuclei of the atoms involved can be considered to be stationary during the transition. This principle, proposed by J. Franck in 1925 and developed on the basis of quantum mechanics by E. U. Condon in 1928, is important in discussing systems where more than one atom is involved. It is therefore valuable in problems of molecular spectroscopy and in the interpretation of the optical properties of liquids and solids.

An example of the principle's operation is luminescent centers in solids. In the state of lowest energy (ground state), the distance from a given center to neighboring luminescent centers adjusts itself so that the total energy of the system

sumes a minimum value. For an ionic crystal, such as sodium chloride, there is a balance between the electrostatic and other forces tending to pull the ions together, and the repulsive forces tending to keep the ions apart. In a state of higher energy (excited state), the system also has a minimum energy position, but this is generally different from that of the ground state because of a different balance of the forces involved. If the luminescent system is in its ground state and absorbs light which raises it to an excited state, the Franck-Condon principle specifies that immediately after the transition the nuclei involved are still in the equilibrium configuration for the ground state. This configuration does not represent the condition of minimum energy for the excited state, and the system rapidly changes to its new minimum, giving up the excess energy as heat. A similar process occurs in the luminescence transition to the ground state, and heat energy is again given off when the luminescent center readjusts to the ground-state equilibrium position. The center has now returned to its initial state. However, it should be noted that the absorption transition was achieved by the action of light energy only, whereas the system returned to the ground state by giving up heat energy twice and emitting the remaining energy as luminescence. As a result, the luminescent emission is of lower energy than the absorbed light, and since the wavelength varies inversely as the energy, the emission is of longer wavelength than is the absorbed light. This general result, known as the Stokes shift in luminescence, is thus explained on the basis of the Franck-Condon principle. *See* LUMINESCENCE; MOLECULAR STRUCTURE AND SPECTRA

[C.C.K.; J.H.S.]

Bibliography: E. U. Condon, The Franck-Condon principle and related topics, *Am. J. Phys.*, 15:365–375, 1947.

Franklinite

A mineral confined almost exclusively to the zinc deposits of Franklin, New Jersey. Franklinite crystallizes in the isometric system with the octahedron the common form. The hardness is 6 (Mohs scale) and the specific gravity is 5.15. The luster is metallic and the color iron black. It may be slightly magnetic. The composition is (Fe,Zn,Mn)(Fe,Mn)$_2$O$_4$ but ... [illegible] ... at Franklin, New Jersey. *See* WILLEMITE; ZINC; ZINCITE.

[C.S.HU.]

Free energy

A term in thermodynamics which in different treatments may designate either of two functions defined in terms of the internal energy E or enthalpy H, and the temperature-entropy product TS.

The function $(E - TS)$ is the Helmholtz free energy and is the function ordinarily meant by free energy in European references. *See* WORK FUNCTION (THERMODYNAMICS). The Gibbs free energy is the function $(H - TS)$. For the Lewis and Randall school of American chemical thermodynamics, this is the function meant by the free energy F. To avoid confusion with the symbol F as applied elsewhere to the Helmholtz free energy, the symbol G has also been used. A recent development has been the introduction of the name free enthalpy, with symbol G, for the Gibbs function.

Theory. For a closed system (no transfer of matter across its boundaries), the work which can be done in a reversible isothermal process is given by

$$W_{rev} = -\Delta A = -\Delta(E - TS) = -(\Delta E - T\,\Delta S)$$

See THERMODYNAMICS (CHEMICAL). For these conditions, $T\,\Delta S$ represents the heat given up to the surroundings. Should the process be exothermal, $T\Delta S < 0$, then actual work done on the surroundings is less than the decrease in the internal energy of the system. The quantity $(\Delta E - T\,\Delta S)$ can then be thought of as a change in free energy, that is, as that part of the internal energy change which can be converted into work under the specified conditions. This then is the origin of the name *free energy*. Such an interpretation of thermodynamic quantities can be misleading, however; for the case in which $T\,\Delta S$ is positive, the above equation shows that the decrease in "free" energy is greater than the decrease in internal energy.

For constant temperature and pressure in a reversible process the decrease in the Gibbs function G for the system again corresponds to a free energy change in the above sense, since it is equal to the work which can be done by the closed system other than that associated with its change in volume ΔV under the given constant pressure P.

$$\Delta G = -(\Delta H - T\,\Delta S) = W_{net} = W_{rev} - P\,\Delta V$$

since

$$\Delta H = \Delta E + P\,\Delta V$$

Each of these free-energy functions is an extensive property of the state of the thermodynamic system. For a specified change in state, both ΔA and ΔG are independent of the path by which the change is accomplished. Only changes in these functions can be measured, not values for a single state.

The thermodynamic criteria for reversibility, irreversibility, and equilibrium for processes in closed systems at constant temperature and pressure are expressed naturally in terms of the function G. For any infinitesimal process at constant temperature and pressure,

$$-dG \geqq \delta w_{net}$$

If δw_{net} is never negative, that is, if the surroundings do no net work on the system, then the change dG must be negative or zero. For a reversible differential process, $-dG = \delta w_{net}$; for an irreversible process, $-dG > \delta w_{net}$. The free energy G thus decreases to a minimum value characteristic of the equilibrium state at the given temperature and

pressure. At equilibrium, $dG = 0$ for any differential process taking place, for example, an infinitesimal change in the degree of completion of a chemical reaction. A parallel role is played by the work function A for conditions of constant temperature and volume. Because temperature and pressure constitute more convenient working variables than temperature and volume, it is the Gibbs free energy which is the more commonly used in thermodynamics.

Partial molal quantities. For a particular homogeneous phase in the absence of surface, gravitational, and magnetic forces, the free energy G depends on the numbers of moles of the constituents present, the temperature T, and the pressure P. Let Ω represent the total number of constituents, n_i the number of moles of typical constituent i, and designate by subscript n constant composition, by subscript n_j constancy of the number of moles of all constituents except n_i. Then

$$dG(T,P,n_1, \ldots ,n_\Omega)$$

$$= \left(\frac{\partial G}{\partial T}\right)_{P,n} dt + \left(\frac{\partial G}{\partial P}\right)_{T,n} + \sum_{i=1}^{\Omega} \left(\frac{\partial G}{\partial n_i}\right)_{T,P,n_j} dn_i$$

Here $\left(\frac{\partial G}{\partial n_i}\right)_{T,P,n_j}$

is the chemical potential μ_i of the ith constituent. It is identical to the partial molal free energy $\bar{G}_i$ of Lewis and Randall (*see* SOLUTION). It then follows that

$$dG = -S\,dT + V\,dP + \sum_{i=1}^{\Omega} \mu_i\,dn_i$$

Because the chemical potentials at constant T,P are intensive variables whose values are fixed, like that of the density, by the relative number of moles of the various constituents present, and are independent of the total mass of the phase, this equation can be integrated for constant T,P and relative composition starting from $n_i = 0$ to obtain

$$G(T,P,n_1, \ldots ,n_\Omega) = \sum_{i=1}^{\Omega} n_i\mu_i$$

This yields

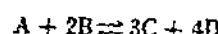
$$dG = \sum_{i=1}^{\Omega} \mu_i\,dn_i + \sum_{i=1}^{\Omega} n_i\,d\mu_i$$

Consistency with the previous expression for dG requires that

$$S\,dT - V\,dP + \sum_{i=1}^{\Omega} n_i\,d\mu_i = 0$$

This is the Gibbs-Duhem equation. For constant temperature and pressure, this relation imposes a condition on the composition variation of the set of chemical potentials.

Heterogeneous systems. The free energy of a closed, heterogeneous system is the sum of the free energies of its various phases. In the absence of such a constraint as provided by the subdivision of the system by a rigid, semipermeable membrane, the general thermodynamic criterion of equilibrium requires that the temperature and pressure be uniform throughout the system and that the chemical potential of each constituent have a common value for all phases in which it is present. Further, if any of the constituents can be formed from others, the chemical potentials of the reactants and products are related in accordance with the stoichiometry of the reaction equation. Thus, for the reaction



$$A + 2B \rightleftharpoons 3C + 4D$$

at equilibrium

$$\mu_A + 2\mu_B = 3\mu_C + 4\mu_D$$

Expressing each chemical potential μ_i in terms of the standard value μ_i^0 and its associated activity term $RT \ln a_i$. See ACTIVITY (THERMODYNAMICS).

$$RT \ln \left(\frac{a_C^3 a_D^4}{a_A a_B^2}\right)_{equil} = -(3\mu_C^0 + 4\mu_D^0 - \mu_A^0 - 2\mu_B^0)$$
$$= -\Delta G^0$$

Here ΔG^0 is called the standard free-energy change for the reaction. Its value depends on the standard states chosen; but for a given temperature and pressure, it is a constant characteristic of the reaction involved. A true equilibrium constant K then results

$$K = \left(\frac{a_C^3 a_D^4}{a_A a_B^2}\right)_{equil} \qquad RT \ln K = -\Delta G^0$$

If the pressure for each standard state is fixed and independent of the pressure of the reaction system, ΔG^0 and hence K are functions of temperature only. This is the conventional approach in the treatment of gas-phase equilibria, but not ordinarily for condensed phases.

Since the activities can be correlated with partial pressures or concentrations through fugacity coefficients or activity coefficients, this thermodynamic approach eliminates the uncertainties otherwise associated with equilibrium calculations based on the law of mass action. *See* EQUILIBRIUM, CHEMICAL; FUGACITY.

The prediction of an equilibrium constant then requires the calculation of ΔG^0 for the reaction. The so-called third-law method involves calculation for the reaction at 25°C of the value of ΔH^0, the standard heat of reaction, from tabulated standard heat of formation data and of ΔS^0 from tabulated third-law entropies. These are combined in the sense of $\Delta G^0 = \Delta H^0 - T\,\Delta S^0$ to permit calculation of the equilibrium constant for 25°C. This in turn is used for evaluation of the integration constant in the integration of the relation

$$\frac{d \ln K}{dT} = \frac{\Delta H^0}{RT^2}$$

The integration requires expression of ΔH^0 as a function of temperature, which necessitates a knowledge of the heat capacities $C^0_{P(i)}$ for the various reactants over the temperature range involved.

Alternatively, if values of the free-energy function

$$\left(\frac{G^0 - H^0_{298}}{T}\right)$$

are available, either from experimental measurement or from statistical thermodynamical computations, they can be combined with the standard heat of reaction at 25°C to give the desired result:

$$\Delta G^0 = \Delta\left(\frac{G^0 - H^0_{298}}{T}\right) + \frac{\Delta H^0_{298}}{T}$$

See ENTROPY; HEAT CAPACITY; THERMOCHEMISTRY. [P.D.]

Bibliography: K. G. Denbigh, *Principles of Chemical Equilibrium*, 1955; J. W. Gibbs, *Collected Works*, vol. 1, 1948; F. T. Wall, *Chemical Thermodynamics*, 1958

Free fall

The ideal unimpeded falling motion of a body acted upon by the pull of the earth's gravitational field. For a body falling near the surface of the earth, the downward acceleration produced by gravity is approximately 32.2 feet per second per second (ft/sec²) or 9.8 meters per second per second (m/sec²). If the distance of fall is relatively small, the acceleration of gravity may be considered constant. For a body falling from rest, the velocity at the end of the first second is 32 ft/sec (approx). At the end of the next second, it is $2 \times 32 = 64$ ft/sec, and at the end of the third second it is $3 \times 32 = 96$ ft/sec, and so on. The average velocity for the first second is 16 ft/sec, and the distance the body falls during the first second is 16 ft. The average velocity for the next second is 48 ft/sec, and the distance of fall is 48 ft. During the first 2 seconds, the distance of fall is $16 + 48 = 64$ ft. If g is the acceleration of gravity, the velocity of a body falling freely from rest after t seconds is

$$v_t = gt$$

The distance of fall h in t seconds is

$$h = \tfrac{1}{2}gt^2$$

and the velocity after freely falling a distance h is

$$v = \sqrt{2gh}$$

For a body falling from a great height, g varies inversely as the square of the distance from the center of the earth. At a height of about 4000 miles above the earth's surface, the distance is twice that from the center of the earth to the surface, and the value of g would be one-fourth its value at the earth's surface.

To compute the velocity attained by a body falling from a height that is large compared to the radius of the earth, the variation of the acceleration of gravity must be considered. Integration of the force equation leads to the result

$$v = \sqrt{2gh\left(\frac{R}{h+R}\right)}$$

In this equation h is the elevation of the point above the surface of the earth at which the object starts to fall. R is the radius of the earth, and g is the acceleration of gravity at the surface of the earth. If the body falls from a very great (infinite) distance, R is negligible with respect to h and the velocity becomes

$$v = \sqrt{2gR} = 7 \text{ mi/sec (nearly)}$$

The accompanying table gives the data for a body falling freely from rest near the earth's surface. *See* GRAVITATION; TERRESTRIAL GRAVITATION. [R.D.RU.]

Data for free fall near earth's surface (approximate)

Time, sec	Acceleration, ft/sec²	Velocity, ft/sec	Distance of fall, ft
0	32	0	0
1	32	32	16
2	32	64	64
3	32	96	144
4	32	128	256
5	32	160	400
10	32	320	1600

Free field

An isotropic homogeneous sound field that is free from all bounding surfaces. The term isotropic implies that the sound propagates in the same manner in all directions, and the term homogeneous implies that the parameters describing the field have the same physical properties at every point. Freedom from bounding surfaces implies that no reflected waves are present in the field. For example, in a free field, a spherical wave propagating outward from a source will always have a spherical wavefront regardless of the distance from the source. *See* SOUND [W.J.G.]

Free radical

The simplest definition of a free radical states that it is any molecule (or atom) which possesses an unpaired electron. Molecules such as nitric oxide and oxygen satisfy this definition, but are not normally considered to be free radicals. Therefore the term should be restricted to exclude ordinary stable molecules, no matter what their valence state may be.

Common examples of free radicals are H, Cl, OH, CN, and CH_3. Such radicals are of great importance in thermal and photochemical reactions, polymerization, and combustion. They are important both in the liquid and the gaseous phase; gas-

phase systems are, however, much simpler, and permit much more decisive interpretation. Even in the gas-phase, experimental methods are inevitably complex and indirect, since substances of such short life cannot be prepared in high concentration. As a result, atoms and radicals must usually be prepared in situ in the presence of a very large excess of other substances. These factors, therefore, make the production, estimation, and detection of radicals a matter of great difficulty and some uncertainty.

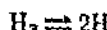
In general, free radicals are formed by the rupture of a bond in a stable molecule with the production of two fragments, each with an unpaired electron. The resulting fragments may disappear in a variety of ways, but especially by their recombination



$$R_1—R_2 \rightleftharpoons R_1 + R_2$$

In most cases recombination occurs at nearly every collision of R_1 and R_2, and the composition of the equilibrium mixture under ordinary circumstances indicates only a minute amount of decomposition into radicals. There are also many ways, other than recombination, by which radicals may disappear. The radicals usually exert partial pressures of less than 10^{-6} mm of mercury and are of short life (usually below 10^{-3} sec). In spite of this they are frequently of great importance in reaction kinetics. The transitory existence of such atoms and radicals has been verified by spectroscopic study.

Production of atoms and radicals. The usual methods of producing free radicals are of three types—thermal, electrical, and photochemical. In thermal methods a stable molecule is decomposed at an elevated temperature. In exceptional circumstances the dissociation into radicals at equilibrium may be considerable. Thus hydrogen atoms may be produced by heating hydrogen to a very high temperature



$$H_2 \rightleftharpoons 2H$$

At 1900°K this equilibrium corresponds to 1% dissociation into atoms when the pressure is 1 atmosphere. In a few cases, substances in solution are considerably dissociated into radicals at room temperature. It is then possible to obtain radicals of apparent long life in high concentrations. Examples are hexaphenylethane, $(C_6H_5)_3C—C(C_6H_5)_3$, which is about 3% dissociated into triphenyl methyl radicals, $C(C_6H_5)_3$, in benzene solution at 5°C at a concentration of 2–3%, and hexa-(*p*-biphenylyl)-ethane,



$$(C_6H_5—C_6H_4)_3C—C(C_6H_5—C_6H_4)_3$$

which is virtually 100% dissociated under similar circumstances. Such radicals owe their special properties to the presence of bulky substituents, which by so-called steric hindrance, slow down the recombination of the radicals (or, stated another way, in which resonance effects weaken the central C—C bond in the parent molecule). Radicals such as triphenyl methyl are essentially ordinary organic substances, with certain distinctive properties as far as addition reactions are concerned.

Usually, however, thermal decompositions are essentially irreversible under the conditions employed. Most gaseous organic substances decompose wholly or in part by a mechanism involving an initial split into radicals. An example is the decomposition of ethane, the first step of which is the reaction

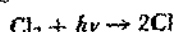


$$C_2H_6 \rightarrow 2CH_3$$

forming two methyl radicals. *See* HYDROCARBON, PYROLYSIS.

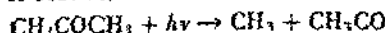
Radicals may also be produced electrically by passing gas through an electrical discharge at high speed. Atomic spectra are produced in this way, and the method is frequently used for the investigation of the chemical reactions of hydrogen, oxygen, and nitrogen atoms. With a fast flow of hydrogen through a discharge tube, for example, it is possible to obtain a gas mixture in which over 50% of the hydrogen is in the form of atoms at low pressures (0.1–1 mm). The method is also useful for the preparation of simple radicals such as OH. It is not of much use for more complex radicals such as CH_3 or C_2H_5, since organic substances are usually decomposed into a wide variety of fragments and the resulting systems are far too complex to be useful.

Photochemical methods of producing free radicals are by far the most common. Almost all gaseous organic compounds decompose photochemically via free radicals, and the method is thus of very wide applicability (*see* PHOTOCHEMISTRY). Two of the most widely used substances are chlorine and acetone. Chlorine decomposes on irradiation with light in the continuous region of its absorption spectrum to give chlorine atoms



$$Cl_2 + h\nu \rightarrow 2Cl$$

Many reactions of chlorine atoms have been investigated by this method. The acetone photolysis has also been widely investigated. There is no doubt that the primary split, using radiation in the range 2537 A to 3130 A, is



$$CH_3COCH_3 + h\nu \rightarrow CH_3 + CH_3CO$$

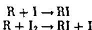
The reaction has been one of the most frequently used sources of methyl and acetyl radicals. Some of the subsequent reaction steps will be mentioned later.

Radiation of short wavelength and high-energy particles, such as occur in nuclear disintegrations, can also lead to the production of radicals, and of ions as well. Such systems are becoming of increasing importance, but are usually complicated and difficult to interpret.

Detection and estimation of atoms and radicals. The earliest methods of detection involved the chemical properties of the radicals. Later and more reliable methods use absorption spectroscopy and mass spectrometry. The methyl radical

first detected by Paneth by the so-called mirror removal method. Radicals were prepared by thermal decomposition in a very fast flow system employing an inert carrier gas. The resulting stream of radicals plus carrier passed down a tube and over a thin "mirror" of lead deposited on the wall. In the presence of the radicals the lead mirror was removed, and it was proved conclusively that this was due to the formation of $Pb(CH_3)_4$, presumably by the reaction of CH_3 radicals with the lead. Many other radicals were detected in this way, but the results were erratic, and the method has been largely superseded.

Another chemical method of detecting radicals uses traps. If iodine is added to a photochemical system, the rapid reactions



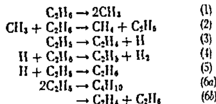
occur. Under appropriate conditions all radicals present may be removed from the system and trapped as iodides. The method has proved very useful in conjunction with other detailed information, but has many pitfalls if used without caution.

Absorption spectroscopy is a simple method of detecting free atoms and has been used successfully for many years. More recently the method has been successfully employed for the detection of OH, NH, NH_2, CN, CF_2, CF, C_2, CH, CHO, C_3, CH_3O, C_2H_5O, NCO, NCS, CH_3S, HNO, PH_2, CH_3, and many other radicals with varying degrees of certainty. The development of flash photolysis, by which large amounts of energy may be absorbed by a gas in a short time, has greatly advanced the detection of radicals by absorption spectroscopy. In an interesting variation of the method, an organic compound in the form of a glass is photodecomposed at very low temperatures, and the radicals formed are thus trapped in a frozen matrix. They have an indefinitely long life, and their spectrum can be determined at will.

The most definite and elegant method for the detection and the estimation of the concentration of radicals is mass spectrometry. A rapid flow system is used in which radicals are produced thermally or photochemically. The gas stream is sampled through the leak of the mass spectrometer. An electron energy is used such that radicals are ionized but stable molecules are not dissociated. Any radical ions detected in the mass spectrometer must come from radicals already present in the gas, and a direct and unequivocal detection of radicals is possible. Also, under favorable circumstances the method can be used for the quantitative estimation of the concentration of radicals.

Finally, in well-investigated photochemical reactions it is frequently possible to establish a mechanism involving a set of free-radical steps beyond any reasonable doubt. When this is possible, the presence and concentration of radicals may be established with a certainty at least as great as that obtainable by any method involving direct estimation.

Free-radical mechanisms. As pointed out above, many reactions, thermal and photochemical, proceed by mechanisms involving a series of free-radical steps. A typical example is the thermal decomposition of ethane, the mechanism of which involves the principal reactions



General mechanisms of this type, the so-called Rice mechanisms, have been well established in many cases by detailed investigation of the kinetics. Several types of free-radical reaction are involved:

1. The rate of the bond-splitting reaction (1) is governed mainly by the dissociation energy of the C—C bond (*see* THERMOCHEMISTRY).

2. The activation energy of the radical decomposition reaction (3) is usually in the range of 10–40 kcal/mole, and the reaction is a typical unimolecular decomposition.

3. Much detailed information on metathesis reactions, such as (2) and (4), has been derived from photochemical investigations. For reactions of this type which involve CH_3, the activation energy is usually in the neighborhood of 7-12 kcal/mole, and the collision theory steric factor is about 10^{-3}–10^{-4}. Results for reactions of C_2H_5 or C_3H_7 are similar (that is, only about one collision in 10^3–10^4 which meet the energy requirements results in chemical reaction). Data for H atom reactions such as (4) are much less certain, but the reactions are much faster than those of methyl radicals. Information on this type of reaction now exists for a large number of reactions of alkyl radicals and, to a limited extent, for radicals containing chlorine, fluorine, oxygen, and nitrogen.

4. Reaction (6*a*) is an example of radical recombination, and (6*b*) of disproportionation. Such radical-radical reactions are rapid and occur at a high percentage of collisions between the reacting species, that is, with a low activation energy and a high steric factor.

The over-all Rice mechanisms are chain reactions, reaction (1) initiating the chain, (3) and (4) propagating it, and (5) and (6) terminating it. Similar mechanisms are involved in a large number of other thermal and photochemical reactions. Gradually a body of information is being built up about the rates of the individual free-radical steps. When such information is widely available, it will become possible to make detailed predictions about the mechanisms of many organic reactions. *See* KINETICS (CHEMICAL). [E.W.R.S.]

Bibliography: W. A. Noyes, Jr. and P. A. Leighton, *The Photochemistry of Gases*, 1941; E. W. R. Steacie, *Atomic and Free Radical Reactions*, 2 vols., 2d ed., 1954; W. A. Waters, *The Chemistry of Free Radicals*, 2d ed., 1948.

free-electron theory of metals

A model of a metal, originally proposed by H. A. Lorentz and improved by A. Sommerfeld in 1928, in which the free electrons, that is, those giving rise to the conductivity, are regarded as moving in a potential—the sum of the average potential due to the metal ions in the lattices and to all the remaining free electrons—which is approximated as constant everywhere inside the metal. The free electrons are thus assumed to move independently of one another throughout the space bounded by the surfaces of the metal. Interaction of the free electrons with the ion cores is assumed to be negligible. The free electrons are taken to be identical to the valence electrons of the free metal atoms; thus, alkali metals contain one free electron per atom, and aluminum contains three. The free-electron theory should actually be termed the quasi-free electron theory, to distinguish between electrons in a metal which behave nearly as do free electrons, and truly free electrons, such as those in a vacuum tube.

A number of important physical properties of some metals, in particular the simple monovalent metals, are explained satisfactorily in terms of the quasi-free electron model. Among these are the electronic specific heat, the magnetic susceptibility, and the electrical and thermal conductivity. To understand the behavior of the more complex metals and of semiconductors, it became necessary to introduce the more complicated band theory of solids. *See* BAND THEORY OF SOLIDS.

Effective mass. In reality, the true crystalline potential seen by a valence electron is not constant, nor independent of the position and motion of the remaining valence electrons. In fact, the potential due to ions located at regular lattice points is periodic in space; in the modern version of the quasi-free electron model the effect of this periodic variation is taken into account by assigning to the conduction electrons an effective mass m^* which is generally different from the mass m of a free electron. The correlation of the motion of a given electron with that of the remaining conduction electrons may be treated theoretically and leads to certain corrections which are of considerable importance in some instances but apparently do not greatly influence the results of the theory of conductivity based on a model which neglects such correlations.

The effective mass m^* is a tensor which relates the force **F** to the acceleration **a** through the equation $\mathbf{F} = m^*\mathbf{a}$. In a coordinate system in which m^* is diagonal the components of the tensor generally are unequal, and some, or all, of the components may be negative. If the components of m^* differ along different crystallographic directions, it is called an anisotropic effective mass. Such anisotropy is the rule rather than the exception and is a consequence of the physical fact that the crystalline potential in which the conduction electrons move is anisotropic. The quasi-free electron model assumes, however, that m^* is a scalar.

Fermi energy. Electrons are elementary particles of half-integral spin; hence they obey Fermi-Dirac statistics. They carry a charge $e = -4.8 \times 10^{-10}$ esu and have an intrinsic magnetic moment μ. According to Fermi-Dirac statistics if there is an assembly of electrons in thermal equilibrium at temperature T, the probability that an allowed quantum state of energy ϵ is occupied is given by the Fermi distribution

$$f_0(\epsilon) = \frac{1}{1 + \exp[(\epsilon - \eta)/kT]} \qquad (1)$$

where k is Boltzmann's constant. The parameter η is called the Fermi energy. It is that energy for which the probability of occupancy is one half. It is also common practice to speak of the assembly of quasi-free electrons as an electron gas. *See* FERMI-DIRAC STATISTICS, KINETIC THEORY OF MATTER.

Wave vector. In quantum mechanics, free electrons are described by wave functions of the form

$$\varphi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \qquad (2)$$

that is, by plane waves. Here **k** is called the wave vector. Its allowed values are determined by the boundary conditions of the problem. For present purposes, **k** may be regarded as a continuous variable. The energy of a free electron is given by

$$\epsilon(\mathbf{k}) = \mathbf{p_k}^2/2m = \hbar^2\mathbf{k}^2/2m \qquad (3)$$

where $\mathbf{p_k}$ is the momentum of the electron in the state **k**, and $\hbar$ is Planck's constant h divided by 2π. The correct wave function for an electron in a crystalline lattice is given not by Eq. (2) but by a function

$$\psi_{\mathbf{k}}(\mathbf{r}) = U_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \qquad (4)$$

known as a Bloch function. Here $U_{\mathbf{k}}(\mathbf{r})$ is a function which is periodic in the lattice. To a first approximation the relationship between the energy and the wave vector is still given by Eq. (3) but with the mass of the free electron replaced by the effective mass m^*.

Density of states. An important quantity which enters into nearly all calculations of electronic properties of metals is the density of states, $N(\epsilon)$; $N(\epsilon)d\epsilon$ is the number of quantum states per unit volume, for a specified electron spin orientation, in the energy range between ϵ and $\epsilon + d\epsilon$. For quasi-free electrons, $N(\epsilon)$ is given by

$$N(\epsilon) = \frac{(2m^*)^{3/2}}{4\pi^2\hbar^3}\,\epsilon^{1/2} \qquad (5)$$

The total number of electrons per unit volume is thus

$$n = 2\int_0^\infty N(\epsilon)f_0(\epsilon)\,d\epsilon = \frac{(2m^*)^{3/2}}{2\pi^2\hbar^3}\int_0^\infty \epsilon^{1/2}f_0(\epsilon)\,d\epsilon \qquad (6)$$

The factor 2 appears in Eq. (6) because each state **k** can accommodate two electrons, the pre-

electrons having opposite spin orientation. The integral in Eq. (6) must be evaluated numerically in the general case. In metals, however, n is so large that $f_0(\epsilon)$ has the form shown in the figure. Nearly all states below the energy η are fully occupied, and the energy range in which $f_0(\epsilon)$ differs significantly from unity and from zero is relatively narrow (of the order kT) compared to the region in which $f_0(\epsilon) = 1$. An electron gas for which this situation holds is said to be highly degenerate. In that case $\partial f_0/\partial\epsilon$ is nonvanishing only over a small energy range, as shown in the figure, and an integral of the form $\int_0^\infty g(\epsilon)\frac{\partial f_0}{\partial\epsilon}\,d\epsilon$ can be expressed in terms of a rapidly converging series:

$$\int_0^\infty g(\epsilon)\frac{\partial f_0}{\partial\epsilon}\,d\epsilon = -g(\eta) - \frac{\pi^2}{6}(kT)^2\left(\frac{\partial^2 g}{\partial\epsilon^2}\right)_{\epsilon=\eta} + \quad (7)$$

The integral in Eq. (6) and similar integrals which occur in the evaluation of properties of quasi-free electrons in metals can be put into the form of Eq. (7) by an integration by parts.

One then finds that

$$\eta = \frac{(2m^*)^{3/2}}{3\pi^2\hbar^3}\,\eta^{3/2}\left[1+\frac{\pi^2}{8}\left(\frac{kT}{\eta}\right)^2 + \quad \cdot\right] \quad (8)$$

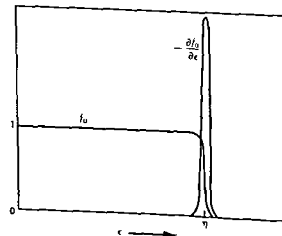
and

$$\eta = \eta_0\left[1-\frac{\pi^2}{12}\left(\frac{kT}{\eta_0}\right)^2 + \quad \cdot\cdot\right] \quad (9)$$

where

$$\eta_0 = \frac{\pi^2\hbar^2}{2m^*}\left(\frac{3\eta}{\pi}\right)^{2/3} \quad (10)$$

is the Fermi energy at $T = 0°K$. The series expansion, Eq. (7), is useful only if $\eta \gg kT$. The dependence of the Fermi energy on the number of electrons per unit volume is apparent from Eq. (10). The fundamental reason for this dependence is found in the Pauli exclusion principle. An



The functions f_0 and $-\partial f_0/\partial\epsilon$ plotted against the energy ϵ. (From F. Seitz, *The Modern Theory of Solids*, McGraw-Hill, 1940)

additional electron which is added to an electron gas that is confined within a fixed volume, will tend to occupy the lowest possible state. If, however, the electrons already present occupy all available states between $\epsilon = 0$ and $\epsilon = \eta_0$, the exclusion principle requires that the additional electrons go into a state whose energy is greater than η_0. Thus as n increases, the energy of that state for which $f_0(\epsilon) = ½$ also increases. *See* EXCLUSION PRINCIPLE.

Electronic specific heat. It is now a relatively simple matter to calculate the equilibrium properties of an electron gas. For example, the average energy per electron is the total energy divided by n, namely

$$\bar{\epsilon} = \frac{2\int\epsilon N(\epsilon)f_0(\epsilon)\,d\epsilon}{\eta} = \frac{3}{5}\eta\left[1+\frac{5}{12}\pi^2\left(\frac{kT}{\eta_0}\right)^2 + \cdot\cdot\right] \quad (11)$$

Equation (11) shows that the contribution to the specific heat per electron, $\partial\bar{\epsilon}/\partial T = C_e$, is given by

$$C_e = k\frac{\pi^2}{2}\frac{kT}{\eta_0} \quad (12)$$

Attention is directed to two features of Eq. (12). First $C_e \ll {}^3\!/_2k$, the classical expression for the specific heat of a free particle. Second, C_e is proportional to the absolute temperature, also in contrast to the classical result. Both of these features are consequences of the degeneracy of the electron gas. Only a small fraction of the electrons in a metal, namely those whose energies are within the small energy range of width equal to approximately $2kT$ and centered at $\epsilon = \eta$, change their energy as a result of thermal excitation. Moreover, the number of electrons which can be excited thermally is proportional to the width of this energy region, and hence proportional to T. Thus, C_e is small, and is proportional to T. *See* SPECIFIC HEAT OF SOLIDS.

The function $f_0(\epsilon)$ gives the equilibrium distribution. Under the influence of applied forces, such as electric and magnetic fields, the steady state distribution $f(\mathbf{k},\mathbf{r})$ of electrons among the allowed energy states differs from f_0. It is convenient to normalize $f(\mathbf{k},\mathbf{r})$ in such a manner that the time average of the number of electrons in the volume element $dxdydzdk_xdk_ydk_z$ is

$$\frac{1}{4\pi^3}f(\mathbf{k},\mathbf{r})\,dxdydzdk_xdk_ydk_z \quad (13)$$

At equilibrium it then follows that

$$f(\mathbf{k},\mathbf{r}) = f_0(\epsilon) \quad (14)$$

Conductivity. In the theory of conductivity, the central problem is that of finding the function $f(\mathbf{k},\mathbf{r})$ in the presence of applied fields and temperature gradients. Once $f(\mathbf{k},\mathbf{r})$ is known, the calculation of electric and thermal currents, for example, is formally trivial. Since the current contributed by one electron is $1ev$, where v is the

velocity of the free electron, the current density due to an electron gas whose distribution is $f(\mathbf{k},\mathbf{r})$ is given by

$$\mathbf{J}(\mathbf{r}) = -\frac{e}{4\pi^3}\int \mathbf{v}_k f(\mathbf{k},\mathbf{r})\, d\mathbf{k} \qquad (15)$$

The thermal current density $\mathbf{Q}(\mathbf{r})$ is obtained by replacing e in Eq. (15) by the energy ϵ; thus

$$\mathbf{Q}(\mathbf{r}) = \frac{1}{4\pi^3}\int \epsilon \mathbf{v}_k f(\mathbf{k},\mathbf{r})\, d\mathbf{k} \qquad (16)$$

Evaluation of Eqs. (15) and (16) leads to expressions for the electrical and thermal conductivities

$$\sigma = \frac{n e^2 \tau}{m^*} \qquad \kappa = \frac{n \pi^2 k^2 \tau T}{3m^*} \qquad (17)$$

Here σ is the electrical conductivity, κ is the thermal conductivity, and τ is called the relaxation time of the electrons. *See* RELAXATION TIME (ELECTRONS); *see also* BOLTZMANN TRANSPORT EQUATION; CONDUCTION (HEAT); THERMIONIC EMISSION. [F.J.B.]

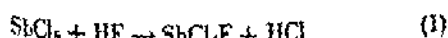
Bibliography: F. Seitz, *The Modern Theory of Solids*, 1940; F. Seitz and D. Turnbull (eds.), *Solid State Physics*, vol. 1, 1955, vol. 4, 1957; A. H. Wilson, *The Theory of Metals*, 2d ed., 1953.

Freon

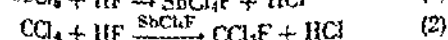
One of a group of polyhalogenated derivatives of methane and ethane containing fluorine and, in most cases, chlorine or bromine. These include such compounds as trichlorofluoromethane (CCl_3F) and dichlorodifluoromethane (CCl_2F_2), also known as Freon-11 and Freon-12, respectively. The freons possess such noteworthy characteristics as nonflammability, excellent chemical and thermal stability, and low toxicity. Additional properties include high density, low boiling point, low viscosity, and low surface tension. This unique combination of properties has made the freons particularly suitable for use as refrigerants.

The freons are usually chemically and thermally stable. This stability results directly from the presence of fluorine atoms in the molecule; generally, the more fluorine present, the greater the stability of the compound. Although freon compounds are related, each member of the series has a different degree of stability and a different chemical structure.

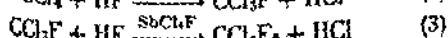
The most popular freons are Freon-11 and Freon-12. These compounds are prepared by the reaction of hydrogen fluoride with carbon tetrachloride in the presence of a catalyst. This catalyst ($SbCl_4F$) is obtained by a reaction between antimony pentachloride and hydrogen fluoride, Eq. (1). Chlorine in carbon tetrachloride is replaced by fluorine, Eqs. (2) and (3).



$$SbCl_5 + HF \rightarrow SbCl_4F + HCl \qquad (1)$$



$$CCl_4 + HF \xrightarrow{SbCl_4F} CCl_3F + HCl \qquad (2)$$



$$CCl_3F + HF \xrightarrow{SbCl_4F} CCl_2F_2 + HCl \qquad (3)$$

Freon-11 (CCl_3F), boiling point 23.8°C (74.8°F), is widely used in commercial and industrial air-conditioning systems and water coolers. Freon-12 (CCl_2F_2), boiling point −29.8°C (−21.6°F), is the most commonly used of the freon refrigerants. Applications include household and commercial refrigeration and air-conditioning.

Other freons are chlorodifluoromethane ($CHClF_2$); chlorotrifluoromethane ($CClF_3$); 1,2-dichloro-1,1,2,2-tetrafluoroethane ($CClF_2$-$CClF_2$); 1,1,2-trichloro-1,2,2 trifluoroethane ($CClF_2$-CCl_2F). These are known as Freon-22, Freon-13, Freon-113, and Freon-114, respectively. Some bromine-containing freons are bromotrifluoromethane ($CBrF_3$) and 1,2-dibromo-1,1,2,2-tetrafluoroethane ($CBrF_2$-$CBrF_2$). These are Freon-13B1, and Freon-114B2.

In addition to their use as refrigerants, the freons serve as propellants in aerosol products, as solvents, and as intermediates in the synthesis of other fluorine compounds. The most common propellant is Freon-12, alone or in combination with another freon. Freon-13B1 ($CBrF_3$) and other freons containing bromine are used as fire-extinguishing agents. Lack of odor, in addition to the properties previously mentioned, render the freons especially desirable for these purposes. As solvents, the freons rank above the hydrocarbons and below chlorinated solvents. Further application has been in the field of polymers and plastics, where several of the freons are important intermediates. *See* AEROSOL; FLUORINE; FLUOROCARBON, HALOGENATED HYDROCARBON; REFRIGERATION. [E.CL.; M.C.C.]

Frequency (wave motion)

The rate at which sound pressure, particle velocity, or other quantities specifying a wave vary from their equilibrium values with respect to time. The most common unit of frequency is the cycle per second (cps). In one cycle, there is a positive variation from equilibrium, a return to equilibrium, then a negative variation and return to equilibrium. This relationship is generally described in terms of the sine wave, and the frequency referred to is that of an equivalent sine-wave variation in the parameter under discussion. *See* SINE WAVE.

Frequency is a convenient means for describing the various ranges of interest in wave motion. For sound waves, the lower and upper limits of the audible frequency range (audio range) are approximately 20–20,000 cps. Frequencies below approximately 20 cps are termed infrasonic, and those above the audible range are termed ultrasonic. Electromagnetic waves vary in frequency from 1 cps for commutated direct current up to 10^{23} cps for the most energetic γ-rays that have been observed. *See* ANGULAR FREQUENCY; FREQUENCY MEASUREMENT; WAVE MOTION. [W.J.G.]

Frequency counter

An electronic device capable of counting the number of cycles in an electrical signal during a preselected time interval. The modern high-sp

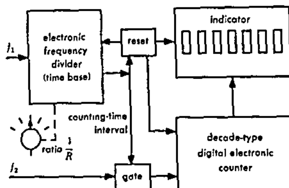


Fig. 1. Basic block diagram of electronic digital counter.

electronic counter is a useful tool in the measurement of frequency when an accurate time base is available. It provides a digital counting or scaling device for registering the total number of events occurring during a given time interval. Such electronic counting circuits operate reliably up to rates of 10^7 counts/sec with reasonably constant input frequencies, such as those produced by electronic oscillators. If it is desired to measure the frequency of an oscillator at 10 Mc or below, the counter may be used directly with a standard frequency oscillator as a time-base control.

Figure 1 shows a block diagram of an electronic digital counter. When f_1 is a standard frequency and $1/R$ is chosen to provide a standard time interval, the counter counts a number of cycles, or pulses, proportional to the frequency of f_2 in cycles per second (or other time interval). When f_1 is not a standard frequency, the counter counts a number proportional to the ratio of f_2 to f_1. When f_1 is a standard fre[illegible]

. .

reading for a harmonic (N) or submultiple ($1/N$) of f_2. When f_1 is a relatively low frequency, that is, its period corresponds to the time interval between two pulses, and f_2 is a standard frequency, the counter may be used to read the period or time interval by setting $R = 1$ and arranging the gate to open on the first pulse and close on the second.

When the frequency to be measured is above 10 Mc, it is necessary to extend the range of the counter. A means commonly used is the heterodyning of the unknown signal by a standard frequency within 10 Mc of the unknown, and subsequent measurement of the beat frequency (see Fig. 2). If the digital counter can count frequencies up to 10 Mc, and the standard-frequency multiples are spaced by 10 Mc, continuous coverage is possible as far up in frequency as the standard frequencies are available. A photograph of a typical electronic digital counter is shown in Fig. 3, with plug-in converter units for range extension.

The principal requirements are for signals of adequate level, free from interference and noise. With these conditions satisfied, it is usually possible to obtain a precision of reading well beyond the stability of the over-all system if a sufficiently long counting time is used. An extension of the frequency-measurement range is made possible by extending the range of the converter or heterodyning system. It is also feasible to add a transfer oscillator (Fig. 4) with harmonics available up into the microwave region. A typical transfer oscillator designed for this purpose is shown in Fig. 5.

The microwave frequencies are then measured as described above (transfer-oscillator method) using the counter to measure the fundamental frequency of the transfer oscillator, and with a beat detector and indicator to determine the beat note between the unknown and a harmonic of the transfer oscillator. A variation of this transfer-oscillator system provides for setting the time-base and zero-set conditions of the counter to give direct reading numbers valid for the particular harmonic of the transfer oscillator in use.

The electronic digital counter is limited to a direct range of approximately 0–10 Mc. Thus it

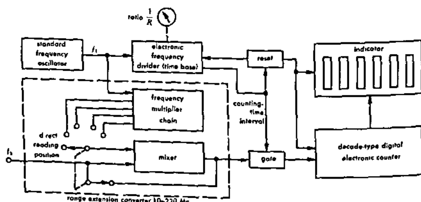


Fig 2. Block diagram of electronic digital frequency counter with heterodyne method of extending useful frequency range.

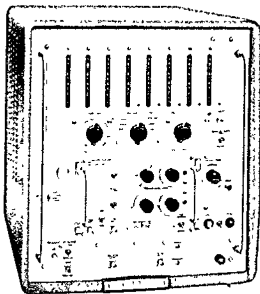


Fig 3. Typical electronic digital counter. Plug-in converter units are included for extending the instrument's range. (*Hewlett-Packard Co.*)

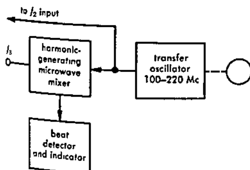


Fig 4 Microwave transfer oscillator.

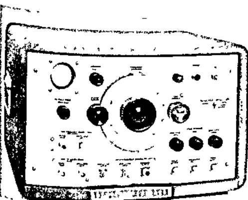


Fig 5 Microwave transfer oscillator designed to be used in conjunction with a digital counter. (*Hewlett-Packard Co.*)

may be considered as an automatic interpolation device for measuring the beat note of an unknown signal heterodyned by a standard-frequency signal, with a maximum range of 10 Mc. All of the usual precautions must be observed to prevent incorrect readings near zero beat with the standard frequency or at points half way between standard frequencies. The ultimate accuracy of the frequency measurement depends on the excellence of the standard frequency oscillator, which is used both for controlling the time base and for providing the known standard frequency harmonics for range extension *See* ELECTRICAL MEASUREMENTS; FREQUENCY MEASUREMENT. [F.D.L.]

Bibliography: F. E. Terman and J. M. Pettit, *Electronic Measurements*, 2d ed., 1952.

Frequency divider

An electronic circuit that produces an output signal at a frequency that is an integral submultiple of the frequency of the input signal. This is required, for example, in television transmission, where a precise relation must be maintained between the horizontal scanning frequency and the vertical scanning frequency.

One technique for frequency division uses a synchronized multivibrator. The original signal, if it is a sine wave, is amplified, clipped, and differentiated to produce pulses which are used to synchronize the astable, or free-running, multivibrator. The output frequency depends upon the natural frequency of the multivibrator but must always be an integral submultiple of the input frequency. If a sinusoidal output signal is required, the square wave from the multivibrator is fed into a tuned amplifier. [illegible]

[illegible] stages and feedback (such as in a decade counter). The output frequency can be any desired integral submultiple of the input frequency. *See* COUNTING CIRCUIT, *see also* SCALING CIRCUIT. [H.F.K.]

Frequency measurement

Measurement of the frequency of a periodic quantity, defined as the number of times a cyclic phenomenon occurs per unit of time. The second is the universally used unit of time. Conversely, time may be measured by observing the number of cycles occurring at constant frequency. The ordinary pendulum and household electric clocks are examples of such time-measuring devices. In the pendulum clock the pendulum completes one cycle of mechanical motion when it returns to its original position from the opposite direction. The electric clock completes one cycle in the electrical sense when the rotor of its motor has turned 360 electrical degrees. The alternating-voltage driving force starts a cycle from zero, passes through a maximum in one direction, returns to zero, passes through a maximum in the opposite direction and

again returns to zero, thereby completing one cycle of 360 electrical degrees. A 2-pole synchronous motor consisting of one pair of north and south poles will turn 360 mechanical degrees for each cycle of 360 electrical degrees.

Standards of time and frequency. The standard second is defined as 1/86 400 of a mean solar day, being an average interval taken by the sun to cross a given meridian. The length of the solar day does not remain constant through the year, the mean solar day being obtained by computing the average taken over a year. The United States Naval Observatory derives the sidereal day from observations of the passage of the fixed stars through the vertical meridian. The sidereal day is used to check and regulate the standard of time.

A universal standard of time consists of some sort of accurate clock. The pendulum-type clock was most reliable in the past, but in recent years piezoelectric-crystal oscillators having constant periods of vibration are used to operate the present-day astronomical clocks which are considered as primary standards at the U.S. Naval Observatory and other locations. If the frequency is determined directly in terms of time, its standard is termed primary frequency standard.

These clocks, also called synchronometers, consist of a quartz-crystal oscillator operating at 100 kilocycles (kc). The crystal oscillator, mounted in a temperature- and pressure-controlled compartment, is made to drive a clock whose rate is determined by comparison with the Naval Observatory time. All over the world the responsibility of establishing and maintaining accurate determination of time is assumed by observatories specially equipped for that purpose. In the United States high-precision radio signals are transmitted by the U.S. Naval Observatory, the facilities being provided by U S. Naval Radio Service. The National Bureau of Standards maintains a standard-frequency broadcasting service over Station WWV, the
Nav
5.0,
beir ... cycles per second (cps).

Primary frequency standard. Figure 1 shows a frequency measurement equipment used as a primary standard. A quartz-crystal oscillator operates at 100 kc and is provided with a frequency-divider chain. The output is divided successively by factors of 10 in order to obtain the fundamental frequencies of 10, 1.0, and 0.1 kc for the three multivibrators. A fourth multivibrator is locked-in directly with the 100-kc signal; hence its frequency is the same as that of the crystal, but its output wave is rich in harmonics at 100-kc intervals for use at high radio frequencies. A cathode-ray oscilloscope is used to compare different known frequencies in the radio- and low-frequency ranges.

A 1-kc multivibrator in the equipment operates a 1000-cps precision clock (syncronometer) in

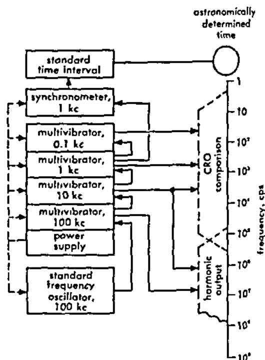


Fig. 1. Block diagram showing the functional arrangement of the primary frequency standard. (*General Radio Co.*)

which a contactor opens and closes each second. Changing the time of contact operation is achieved by turning an adjustable dial with 100 scale divisions. This contactor is adjustable in phase, or time of operation with respect to an arbitrary zero, from 0 to 999 milliseconds (msec) and around through zero, into the next second. A calibrated control is provided and is graduated in 10-msec increments from 0 to 100 in 360° or 0 to 1000 msec. Periodically the syncronometer time is checked against astronomically determined time, and the operation of the contactor is adjusted to correct the syncronometer operation.

Secondary standard. The frequency of a secondary standard is determined by comparison with a primary standard, or by comparison with another secondary standard originally compared with a primary standard. The equipment for a secondary standard is exactly the same as for a primary standard except that it does not use an electric clock.

Checking or adjusting the accuracy of a secondary standard can be achieved by comparing a radio-frequency wave of known frequency with one of the harmonics in a 100-kc multivibrator signal. For example, if Station WWV is sending out a 5-Mc signal, the 100-kc multivibrator signal would be coupled into a radio receiver's antenna and a beat between the 5-Mc signal and the 50th harmonic of the multivibrator signal. If the beat occurs in 2 sec, the difference in frequency is one part in 10^7. Such a slow beat usually can be

measured by observing the swing of the pointer on a dc indicating instrument.

Audio-frequency meters. A fairly wide variety of frequency meters has been developed and is commercially available. Broadly speaking they can be grouped into two classes, the resonant and ratiometer types. The resonant type may be further subdivided into instruments employing resonant reeds and those having electrically resonant circuits. The latter, as well as the ratiometer type, are classified as deflection-type instruments embodying moving systems with pointers and scales.

The moving systems of the conventional deflection-type meters develop two opposing deflecting forces which cause the movement to come to rest when the two opposing torques are balanced. The net torque is directly dependent on the frequency being measured. Deflection-type frequency meters, commonly used in power-frequency applications around 25, 50, and 60 cps, are also available in other ranges up to about 900 cps. Their accuracies may be in the order of 0.14 cps in the low-frequency ranges and about 4 cps in the higher frequency ranges.

Resonant-reed-type frequency meters are available in various ranges between 10 and 1000 cps, with special designs ranging as low as about 7 cps and as high as about 1500 cps. Their accuracies, which are independent of their frequency ranges, can be in the order of ±0.1% of specified frequency.

Reed-type frequency meters. This type of meter uses the principle of vibrating reeds. The reeds for such instruments, assembled into a so-called reed comb, are made of specially selected and properly tempered steel. They have bent tips which are enameled white for visibility. When the supply voltage is applied to the instrument, all of the reeds receive vibrational impulses; the effect is visible only in the reed or reeds which are in resonance. The reed in resonance behaves somewhat like a whip, the tip swinging through a readily visible arc, the natural period of vibration depending upon length and thickness. Resonance vibration extends over a limited visible range from about 2% below to 2% above the actual frequency as shown in Fig. 2. Individual reeds can be tuned to a range of about ±0.1% of their rated frequencies.

The reeds in the comb are vibrated by an electromagnet energized from the source of which the frequency is to be measured. Depending upon design considerations, a permanent magnet may be used in conjunction with soft iron in the magnetic circuit. The magnetic-drive system usually takes one of two forms. The direct-drive method drives the reeds magnetically by including them in the magnetic circuit as shown in Fig. 2*a*. In the indirect-drive method, the magnetic circuit includes a soft-iron pole piece or armature attached to the base of the reed comb, the armature indirectly transmitting its mechanical vibrations to the reeds. If no permanent magnet is used in the magnetic circuit, the reeds are vibrated at a frequency twice that of the source of voltage and consequent magnetic field. Biasing the magnetic field by use of a permanent magnet causes the reeds to vibrate at the same frequency as the source voltage. Figure 3

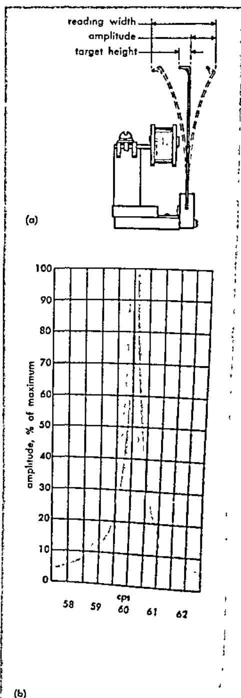


Fig. 2. Reed-type frequency meter. (a) Resonance curve for 60-cps reed showing amplitude of vibration for various frequencies applied (b) Action of a Frahm frequency meter.

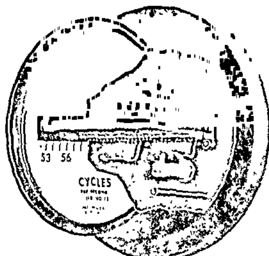


Fig 3. Cutaway view of an indirect-drive, switchboard-type frequency meter. (*James G. Biddle Co.*)

... and B, are placed at right angles to form a *moving element* and are supplied through a resonant circuit. The coils are tuned to different frequencies, *coil A* to a frequency slightly below the lowest scale point and coil B an equal amount above *the upper* scale point. A fixed coil C divided into two parts carries the sum of the currents of the two *circuits of coils A* and B. When the frequency of the applied voltage equals the mid-scale frequency the currents in the ...

... the fixed-coil current. At frequencies near the low end of the scale, the moving element comes to rest with coil A approaching a position parallel to the fixed coil. As *the circuit* of coil A approaches resonance, the ...

... frequencies in the higher range, coil B approaches resonance and *takes* an equilibrium position parallel to the plane of the fixed coil. If the supply frequency is *very low, the* impedance of the circuit A and B being equal, the pointer will indicate a fictitious mid-scale reading. *To rectify* this misleading measurement, a third coil D is introduced in parallel *with coil* A and resonates at frequencies much less than the lowest scale point.

Figure 5 shows a *pictorial* diagram of another moving-coil resonant-type frequency meter based on the electrodynamometer principle and usually called a mutual-ind... field st... which ... ing. This construction provides a *scale* length of approximately 250° and covers a range of 56–65 cps. An alternating flux produced by the field coil induces an electromotive force in the moving coil which varies with the deflection. This emf is opposed by the voltage drop across the tuned circuit. The resultant voltage causes a current to pass through the moving coil. A torque is developed by the interaction between the moving-coil current and the field flux and deflects the coil until it reaches a position where the in-phase component of its current becomes zero, which causes the torque also to become zero. The equilibrium position changes with frequency because the drop across the tuned circuit and the induced emf both vary with frequency.

Moving-iron meters. A resonant type of frequency meter operating on the moving-iron principle is shown in Fig. 6. Two field coils are mounted opposite to each other and are connected so that their fluxes oppose. If i_1 and i_2 are moving-coil currents, the resultant field current will be $i_1 - i_2$. The moving system which lies between the field coils is comprised of an iron vane of magnetic material centered within the armature coils and rigidly attached to the shaft. Both the field coils are connected to series-resonance circuits in which resonance is produced below the operating range for one circuit and above the operating range for the other. Thus for a frequency range of 55–65 cps, the resonance may occur around 50 cps for one circuit and 70 cps for the other.

Deflecting torque is produced by the reaction of the armature, or moving coil, field with the in-phase component of the field coils. This torque is proportional to the product of the resultant field flux,

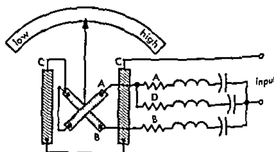


Fig. 4. Resonant-type moving-coil frequency meter.

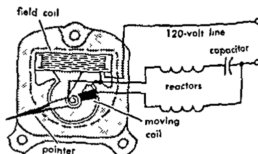


Fig. 5. Mutual-inductance type of frequency meter (*General Electric Co.*)

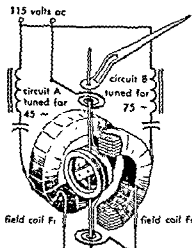


Fig. 6. Moving-iron resonant-type frequency meter. (General Electric Co.)

armature flux, and the cosine of the phase angle of the two fluxes.

In addition to this deflecting torque, there is a countertorque produced by the action of the field flux in the magnetic vane of the moving element. When the deflecting torque causes the iron vane on the moving system to be out of alignment with the field flux, the iron vane exerts a torque opposite in direction (to the deflecting torque) because of its tendency to realign itself with the physical direction of the field flux. This countertorque is required to establish a position of equilibrium. This occurs when the magnitude of the countertorque is equal to the deflecting torque, which causes the pointer to take a position indicating the frequency.

Another design of a moving-iron frequency meter, but of the ratiometer type, is shown in Fig. 7. Two coils A and B are arranged at right angles to each other. Coil A is in series with a resistance R_A and this combination is in parallel with an inductance L_A. Similarly, coil B is in series with L_B and the combination is in parallel with R_B. The complete circuit thus acts as a bridge network. When the bridge is balanced by suitable values of parameters, the currents in the two coils are equal and the moving element, a pivoted soft-iron needle, will indicate the mean-frequency position as shown in the diagram. When the frequency increases the current increases in coil A and decreases in coil B. The interaction of the fields of the two coils will cause the needle to take up a new position. Hence for each change in frequency the needle tends to align itself along the resultant magnetic fields of the two coils. In this kind of meter the effects of distorted waveforms due to higher harmonics are almost nonexistent.

Audio-frequency bridge methods. An ac bridge is generally made up of a four-arm network, where an ac source is applied to a pair of opposite terminals and a current-detecting device is connected to the remaining two terminals. With the bridge network energized, the current in the detecting device can be made zero by adjusting suitable values of resistance, capacitance, or inductance in the four bridge arms, a process known as balancing.

Bridge networks in which the balancing action depends upon the supply frequency may be used for frequency measurement. The choice of frequency bridge is dependent on the frequency range, the available apparatus, and the ease with which the bridge can be set up and balanced. *See* BRIDGE CIRCUIT.

A simple form of resonant-frequency bridge is shown in Fig. 8. Balance is obtained by adjustment of R_4 and C, the detector D showing zero current. This will occur when $R_1/R_2 = R_4/R_3$ and $\omega L = 1/\omega C$ or $\omega^2 = 1/LC$. From this, the frequency is found to be

$$f = 1/2\pi\sqrt{LC}$$

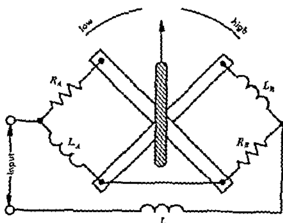


Fig. 7. Moving-iron ratiometer-type Weston frequency meter.

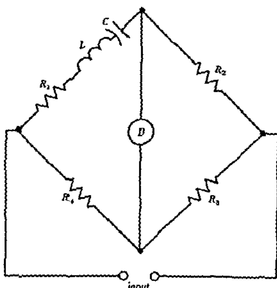


Fig. 8. Resonance-frequency bridge.

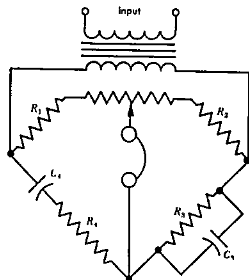


Fig. 9. Wien frequency bridge.

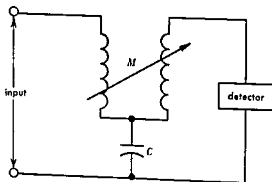


Fig 10. Campbell's frequency bridge

The most commonly employed bridge is the Wien bridge, because it has many advantages of other bridge methods. Figure 9 is the schematic diagram. R_3 and R_4 can be varied together. Condition of balance will be obtained when $R_3 = R_4$, $C_3 = C_4$, and $R_1 = 2R_2$.

The frequency of the source is found to be

$$f = 1/2\pi R_3 C_4$$

If the frequency to be measured is in the audio range, a pair of headphones may be used as a detector; for frequencies beyond the audio range an electronic voltmeter can be used.

Campbell's bridge is shown in Fig. 10. This bridge is used because of its simplicity and fairly large range of frequency. Balance is obtained by adjusting the value of M so that the detector current is zero. Under this condition the voltage induced in the secondary of the mutual inductor M is equal in magnitude and 180° out of phase with the voltage drop across the capacitance C.

Hence at balance $\omega Mi = i/\omega C$ or $\omega^2 = 1/MC$, from which the frequency is shown to be

$$f = 1/2\pi\sqrt{MC}$$

Direct-comparison methods. The most widely used of the simpler methods is aural comparison. An unknown frequency can be determined by direct comparison with a known adjustable standard frequency by using the ear. Figure 11 shows a typical scheme where headphones are connected directly to the two audio sources. The unknown frequency is fed into one earpiece while the standard frequency supplied by a resistance-tuned or a beat-frequency oscillator is fed into another earpiece. A beat note between the two frequencies can be detected by the ear when the frequencies are not exactly equal.

A meter or an electric-eye indicator can take the place of the ear as a detector. An indicating ac voltmeter with a fast d'Arsonval movement may be used. Under the influence of the beat note the pointer of the voltmeter vibrates about a mean value of the output voltage. For low values of frequency difference it swings back and forth; for high values it vibrates rapidly. Such a meter or an electric-eye indicator can detect frequency differences of very low values accurately, whereas in the ear method of detection the accuracy of adjustment is limited by the skill of the operator.

Cathode-ray oscilloscopes provide a most convenient means of comparing two frequencies. The usual method consists of applying a voltage of unknown frequency to one pair of deflecting plates and a voltage of known frequency to the other. The resulting patterns, known as Lissajous figures, are shown in Fig. 12. These patterns are a function of the difference in amplitude, frequency, phase, and the waveforms of the two applied voltages and their frequency ratio. Knowing one of the frequencies and determining the difference or the ratio of the frequencies, it is possible to calculate the second frequency. For example, the pattern with five vertical loops is produced by signals whose frequency ratio is 5:3 with the lower frequency on the horizontal plates. Any value of frequency in the audio range can be determined by this method, providing the ratio of the two frequencies does not exceed an approximate value of 10:1. *See* LISSAJOUS FIGURES.

Frequency counters. For the basic measurement of audio frequencies one of the simplest methods is provided by frequency counters. In the simplest form, a mechanical counter is actuated by a polarized relay through an escapement or a linkage mechanism. The relay operates on each cycle of low-frequency voltage. The frequency values within the operating range can be determined by counting the number of cycles during a known period of time. Such counters are suitable only for very low frequencies; they are not satisfactory for frequencies even as high as 60 cps.

For higher frequencies, electronic counters are useful. A block diagram of a commercial design is shown in Fig. 13. It shows a block diagram of an electronic counter, where the four stages consist essentially of so-called flip-flop circuits. Stages A,

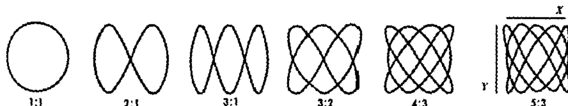


Fig 11. Scheme for aural method of frequency measurement

B, C, and D represent 1, 2, 4, and 8 counts, respectively. All the stages can be in two possible states, 1 or 2, signifying OFF and ON position of their neon lamps.

Starting with all the stages in state 1, when a pulse is sent to the first stage A, it triggers into state 2 and its neon lamp starts glowing. A second pulse sent to stage A will put it back into state 1; however, the pulse will be directed to stage B and its lamp will glow. Stage B is now in state 2 signifying a count of 2. A third pulse will put stage A again into state 2 and now both the lamps of stages A and B will glow, indicating a count of three $(1+2)$. A fourth pulse triggers both the stages A and B into state 1. A pulse is sent to stage C which triggers into state 2. The lamp of stage C starts glowing, representing a count of four. The next pulse will put stage A into state 2 The lamp for stage A and C will glow, indicating a count of five $(1+4)$. The total count in all cases is equal to the sum of the counts represented by the stages which glow. Thus, if three stages A, B, and C glow, seven pulses will be received, indicating seven counts. Usually four-stage counters of this type are designed so that after nine pulses are received a pulse will be sent to the next counter.

In this type of counter, the usual frequency range is 10 cps to 100 kc; however, in some cases, counting rates of 10^6 cps (1 Mc) are possible. *See* COUNTER, DIGITAL; FREQUENCY COUNTER. [E.B.C.]

Radio-frequency measurements. Precise measurement of frequencies above the audible range may be performed by various techniques. Basic measuring systems may consist of (1) a calibrated oscillator with some means of comparison with the unknown frequency, such as an oscilloscope or a heterodyne detector unit, (2) a digital counting or scaling device which registers the total number of events occurring during a given time interval, or (3) an electronic circuit for producing a direct current proportional to the frequency of its input signal which then may be indicated by a dc meter. In practice, each of these basic systems has an upper frequency limit for practical operation.

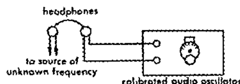


Fig. 12. Lissajous figures for various frequency ratios and phase angles.

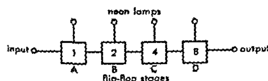


Fig. 13. Block diagram of electric counter

In order to extend the range of frequency measurement above the practical limits mentioned above, it is customary to generate a fixed standard frequency and to select a harmonic of it near the unknown frequency, after which the unknown frequency may be reduced to a lower value by subtraction of the standard-frequency harmonic in a mixer or heterodyne detector. This lower beat frequency may then be measured by application of the basic methods outlined above, which thus serve to interpolate between the known standard frequencies.

The complexity of the equipment required for a given frequency measurement is usually related to the nature of the unknown signal, the amount of manual operation of the controls required, and the precision required. If the unknown frequency is a pure sine wave of local origin or locally controllable, of adequate amplitude, and free from interfering signals, it is possible to measure its frequency by using (1) a digital counter with appropriate range-extending converter and frequency standard, (2) a heterodyne detection and interpolation system with a suitable frequency standard, (3) a calibrated oscillator and detector unit, called a heterodyne frequency meter, or (4) a calibrated absorption circuit such as a wavemeter or grid-dip meter if slightly lower precision is adequate (*see* WAVEMETER). If the unknown frequency is subject to interference, intermittent in nature, is too low in amplitude to override noise, or otherwise requires selective filtering and identification, the simple systems outlined above require additional equipment and manipulation to produce acceptable measurements Additional knowledge and skill are required of the operator. Measurements to high orders of precision require longer time intervals than less precise measurements

Microwave frequency measurements. The measurement of most electrical quantities becomes more difficult as the frequency becomes higher.

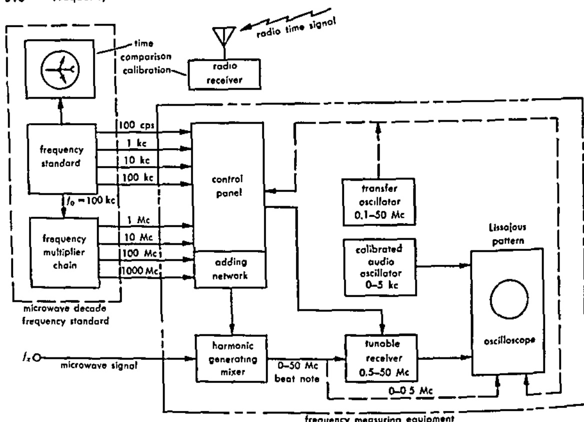


Fig. 14. Block diagram of microwave frequency standard and frequency-measuring equipment.

Fortunately, the measurement of frequency, even in the microwave region, is not of unusual difficulty.

In the microwave range of frequencies, the wavelengths involved are so short that it is easy to measure the wavelength and thus obtain the frequency by a simple calculation Many absorption-type wavemeters are calibrated directly in frequency. The principal difference in technique at microwave frequencies is concerned with the generation of usable harmonics of fixed standard-frequency oscillators or of tunable transfer oscillators in the range where measurements are required. Otherwise, the use of either the digital counter or heterodyne system interpolation between standard-frequency harmonics is carried out as at lower radio frequencies.

If a microwave signal is locally controllable and of a reasonably steady, strong, and stable nature, it is possible to measure it directly by heterodyning it against a known standard frequency and measuring the resulting beat frequency. It is necessary to have a frequency standard, harmonic-generating multiplier chain, mixer, and either a heterodyne system or digital counter system for measuring the beat frequency. Added complexities requiring tunable receivers for selecting the beat frequency and tunable local oscillators for measuring the beat note are sometimes necessary if the beat frequency is not strong. A block diagram of a decade frequency standard and measuring system is shown in Fig. 14.

Transfer oscillator methods. Calibrated oscillators may be used for interpolation between standard-frequency harmonics by using their dial calibrations and conventional interpolation methods. Microwave frequencies are usually measured by setting the frequency of the transfer oscillator so that a harmonic of it is at zero beat with the unknown frequency. In this way, it is possible to use the tunable oscillator to transfer the signal to a lower frequency range where it is easy to measure against a frequency standard by direct means. It is usual to provide a tuning range slightly greater than 2:1 for such a transfer oscillator in order to give continuous harmonic coverage at high frequencies.

Transfer oscillators require both high stability and high harmonic output, conditions which are not mutually compatible in one circuit. The stability is generally obtained by using a low-power oscillator with high-Q tuned circuit, the harmonics by separate harmonic generators.

The development of controlled-frequency or synthesizer-type generators enables generation of accurately known stabilized frequencies, and hence makes possible directly calibrated transfer oscillators or self-calibrated heterodyne frequency meters. The precision and accuracy of a frequency synthesizer depend on several factors, the limiting item ultimately being the frequency standard used.

Heterodyne frequency meters. The calibrated tunable oscillators mentioned above for use as transfer oscillators may be used directly for fre-

quency measurement by their dial calibrations alone. *A beat detector and indicator are necessary,* as well as a harmonic generator covering the required frequency range; a diode mixer and suitable amplifier with headphones or meter indicator may be used. Unless a synthesizer type of stabilized controlled oscillator is used, the accuracy will not be as good as that obtainable by either the direct measurement or the transfer-oscillator method. Accuracies of 0.1% can be obtained with a well-designed portable heterodyne frequency meter. If an internal calibrating crystal oscillator is used, *better accuracy may be obtained. The synthesizer-*type instrument affords accuracy approaching that of its reference frequency standard. [F.D.L.]

Bibliography: H. Buckingham and E. M. Price, *Principles of Electrical Measurements,* 1955; A. Campbell and E. C. Childs, *The Measurement of Inductance, Capacitance and Frequency,* 1935; E. L. Ginzton, *Microwave Measurements,* 1957; F. K. Harris, *Electrical Measurements,* 1952; I. F. Kinnard et al., *Applied Electrical Measurements,* 1956; C. G. Montgomery, *Technique of Microwave Measurements,* 1947; F. E. Terman and J. M. Pettit, *Electronic Measurements,* 2d. ed., 1952.

Frequency modulation

A special kind of angle modulation in which the instantaneous frequency of a sine-wave carrier is varied by an amount proportional to the magnitude of the modulating wave. In many so-called FM applications, the angle modulation is neither FM nor phase modulation (PM) but one in which the angle changes in some other way in accordance with the modulating wave. See ANGLE MODULATION; PHASE MODULATION.

Either AM or FM offers a solution to the important problem of how to impress the message wave to be communicated upon a high-frequency oscillation. Also, either AM or FM permits detection and faithful reproduction of the original message. However, unlike AM, FM does more than just enable communication. FM offers additional *important advantages in exchange for extra band-*width occupancy. Also, FM with negative feedback minimizes noise problems and receiver distortion. See MODULATION.

Instantaneous frequency. Frequency modulation is defined in terms of a generalized concept known as instantaneous frequency which is directly proportional to the time rate of change of the angle of a sine function, the argument of which is a function of time. When the argument is expressed in radians and the time in seconds, the instantaneous frequency in cycles per second is the time rate of change of the angle divided by 2π.

In frequency modulation the instantaneous frequency is linearly proportional to the magnitude of the modulating wave.

Principles of FM. The illustration depicts typical waveforms of AM and FM for increasing magnitudes of a sine-wave modulating signal. Louder tones with AM mean greater changes in amplitude. Louder tones with FM mean greater *changes in frequency.*

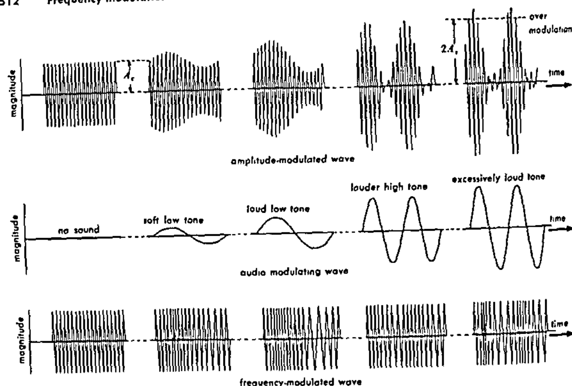
In AM, as the audio volume increases, the peak power of the modulated wave increases, the average power of each sideband increases, the carrier power is unchanged, and the zero crossings (zeros) of the modulated wave are unchanged and correspond to those of the unmodulated carrier, provided the modulation is short of being complete. See AMPLITUDE MODULATION.

In FM, as in any kind of angle modulation, the variable zero crossings of the modulated wave *carry the information of the message to be com-*municated. Neither peak nor average power of the modulated wave changes. Even though the total average power remains constant, its distribution with frequency in the transmitted band changes continuously and in a nonlinear manner as a function of the modulating wave.

J. R. Carson in 1922 was the first to present a mathematical analysis demonstrating the wide band of frequencies involved in FM. However, E. H. Armstrong, in his patent of 1933, was the first to have a real *appreciation of the noise-reducing properties* of FM. At first, there was considerable skepticism about the practical utility of FM. Today, its vast *multitude of uses and widespread acceptance are* ample proof that for some purposes FM is better than either AM or single-sideband modulation (SSB). *Single-sideband* modulation conserves bandwidth, whereas FM uses extra bandwidth to reduce noise still lower.

Advantages of FM. In radio broadcasting, provided the frequency deviation (peak difference between instantaneous and carrier frequencies) is large and provided multipath transmission effects are small, FM is capable of high-fidelity reception combined with the advantages of reduced noise, less interference between stations, and less transmitter power to cover a given area. Constant average power and constant peak power that is only twice the average power are two factors that permit a ready realization of a simple high-efficiency transmitter, *simplify problems of automatic volume con-*trol, and allow amplifiers and other devices to operate closer to their maximum power capability *without the penalties normally associated with* nonlinearities.

Spectrum of FM. The spectrum of any angle-modulated wave extends above and below the carrier frequency by amounts which theoretically extend indefinitely. For example, suppose the modulating signal is a single-frequency tone, 1000 cycles per second (cps) and the carrier frequency is 1,000,000 cps. The lower side-frequencies in cps will be 999,000, 998,000, 997,000, and so on, corresponding to the difference between the frequency of the carrier and the frequency of the modulating signal and its harmonics. If C and V are the carrier and modulating-signal frequencies, the lower side-frequencies are $C - V$, $C - 2V$, $C - 3V$, and so on. The upper side-frequencies in cps are 1,001,000, 1,002,000, 1,003,000, and so on, corr



Waveforms of amplitude-modulated carriers and frequency-modulated carriers with varying degrees of modulation. (*From H. S. Black, Modulation Theory, Van Nostrand, 1953*)

sponding to $C + V$, $C + 2V$, $C + 3V$, and so on. In addition, the carrier amplitude is reduced when the modulating tone is applied and may even become zero.

This example might suggest that the side-frequency components of an FM wave would always be symmetrical about the carrier, independent of the waveform of the modulating wave. Such is not the case. Unlike the symmetrical spectra of the upper and lower sidebands in AM, in FM unsymmetrical side-frequency spectra are a necessary consequence of an asymmetrical modulating wave.

Practically, when estimating approximate bandwidth occupancy, a rule of thumb states that angle modulation requires the band traversed by the instantaneous frequency plus the bandwidth of the modulating wave added at both top and bottom. For some purposes an even wider band may be required.

Moreover, unambiguous representation and recovery of the wanted message by angle modulation techniques also require that the unmodulated carrier frequency comfortably exceed the sum of the frequency deviation in the down direction plus the bandwidth of the modulating wave. In other words, in FM the carrier frequency must be high compared to the maximum frequency deviation.

Noise advantage of FM. For certain types of noise disturbance characterized by a noise spectrum that is uncorrelated and independent of frequency, the ratio of average signal power to average noise power in the output of the FM receiver will be proportional to the square of the peak-frequency deviation. Therefore under certain important conditions, the signal-to-noise ratio of an angle-modulation system improves 6 decibels for each 2:1 increase in bandwidth occupancy.

However, the noise advantage of FM cannot be increased indefinitely. As the bandwidth occupancy is continually increased to accommodate an increased frequency deviation, more noise reaches the FM detector. Presently, the assumption that the noise is less than the so-called improvement threshold is violated, whereupon the noise advantage of FM is quickly lost.

FM with negative feedback acts differently in that the improvement threshold is minimized and held constant, independent of the bandwidth occupied by the incoming FM signal.

Improvement threshold. Any system that reduces noise in exchange for extra bandwidth occupancy is characterized by a threshold effect which becomes pronounced when the reduction is large. For a wide-band FM system this threshold is a critical function of the signal-to-noise ratio at the receiver, after selection and before any nonlinear process such as amplitude limiting. When noise perturbations appreciably exceed this critical value, which is termed improvement threshold, the signal-to-noise ratio at the output of the FM receiver rapidly deteriorates. As the noise is still further increased, noise abruptly grabs control and the desired signal tends to be suppressed from the output.

FM with negative feedback. This is accompanied by a decrease in distortion originating within the FM receiver and by an increased tolerance to

noise falling within the frequency band occupied by the incoming FM signal. These two important advantages are not possessed by nonfeedback receivers. Substantial benefits are realized only when the amount of negative feedback is large. In common with nonfeedback FM systems, any large reduction in noise must be paid for by a corresponding increase in the bandwidth occupied by the transmitted FM signal.

FM with negative feedback provides an efficient method for the detection and tracking of narrow-band signals in the presence of wide-band noise, especially when the receiver must operate in the presence of large and continuously varying Doppler frequency effects. In numerous other situations where the saving of power is vital and where extra bandwidth is available, FM with negative feedback may prove attractive as, for example, in communications by means of satellites.

When the FM signal is carrying a number of communication channels as in a high-grade multiplex system, nonlinear distortion must be extremely small. This extraordinary degree of linearity is readily obtained with negative feedback. An FM receiver with negative feedback is also characterized by improved stability. With enough feedback, the output becomes virtually independent of such factors as fading of the input signal, power supply voltages, or changes in receiver amplification or detector efficiency.

In a typical nonfeedback receiver, the incoming FM signal is shifted to an intermediate frequency (i-f) by a product modulator fed by a beating oscillator. The instantaneous frequency variations of the i-f signal are then detected, thereby reproducing the original audio (modulating) signal.

With feedback, a portion of the receiver output is fed back and caused to modulate the frequency of the beating oscillator. The receiver output is increased or decreased according to whether the feedback is positive or negative. The amount of feedback is measured by the decibel change in output due to feedback. Negative feedback reduces the instantaneous frequency difference between the incoming FM signal and beating oscillator, thereby reducing the frequency swing (degree of modulation) of the i-f signal.

With sufficient negative feedback, the degree of modulation of the i-f signal is reduced to near zero, thereby reducing the i-f bandwidth to little more than twice the top audio frequency. By this means, the i-f signal is separated from most of the noise existing in the wider bandwidth occupied by the incoming FM signal. Under these conditions, regardless of the bandwidth occupied by the incoming FM signal, the transmitter power need only be enough to override noise in the narrow i-f band.

Production and detection of FM. Many schemes are possible and nearly all use spectrum translation. For the production of FM most schemes resort to spectrum multiplication.

Spectrum translation. Spectrum translation of an angle-modulated wave is accomplished by single-sideband modulation. The translated spectrum, with or without inversion, is centered about a new carrier. Otherwise its significant properties are unchanged.

Spectrum multiplication. This implies angle multiplication. By generating the xth harmonic of an angle-modulated wave, the angle is multiplied by x. If the required multiplication is too great, then after a convenient number of multiplications the resulting spectrum may be translated downward and the multiplication resumed.

Angle multiplication by x multiplies spectrum parameters by x. Typical parameters are significant band-width occupancy, carrier frequency, initial carrier phase, instantaneous frequency, instantaneous-frequency deviation, instantaneous-phase deviation, peak-frequency deviation, and peak phase deviation.

Preemphasis and deemphasis. Widespread use is made of the fundamental principle that by linearly distorting (equalizing) the modulating wave in a predetermined manner (thereby in effect creating a new modulating wave) any given angle modulator may be converted to another angle modulator having an arbitrary response. Similarly, appropriate equalization of the filtered output of a particular angle detector permits realization of an arbitrary detector response (see PHASE MODULATION). In radio broadcasting, these procedures are termed preemphasis and deemphasis and are used to improve signal-to-noise ratio.

FM modulators. For the most part FM modulators fall into two classes. In Class I, a phase modulator is converted to a frequency modulator by first passing the audio signal through a circuit whose output is inversely proportional to frequency. The FM wave thus created normally goes to a spectrum multiplier to increase its frequency deviation. In Armstrong's experiments, this phase modulator consisted of a suppressed-carrier amplitude modulator. Carrier was not only supplied to this balanced modulator in the usual way but in addition was shifted in phase by 90° and then added to the modulator output. When this added carrier is sufficiently large, the resultant output approximates the output of a phase modulator. Armstrong intentionally limited the peak phase modulation resulting from this first step to a maximum range of only 30° in order to ensure reasonable linearity. Consequently, the phase had to be multiplied by several thousand by a succession of spectrum multipliers in order to produce a frequency swing of ±75,000 cps with peak audio input. Amplitude limiting in the first multiplier attenuated any residual amplitude modulation associated with the phase-modulation process.

In Class II, a capacitance, inductance, potential of a control electrode, or other parameter is varied directly to change an oscillator frequency. Class II modulators usually require spectrum multiplication. See FREQUENCY MODULATOR.

FM detectors. Ideally, these would p... out ... proportional to changes in ins...

frequency, would not respond to amplitude modulation, and would limit output band to that allocated to the wanted message. In a practical receiver, frequency selection (filtering) is usually followed by spectrum translation and i-f amplification. Finally, frequency detection is usually accomplished by an amplitude limiter followed by a balanced discriminator, followed by a low-pass filter. The discriminator converts frequency variation to amplitude variation. *See* FREQUENCY-MODULATION DETECTOR. [H.S.BL.]

Bibliography: A. N Goldsmith, et al. (eds.), *Frequency Modulation*, Radio Corp. of America, vol. 1, 1948.

Frequency modulator

A circuit or device producing frequency modulation. Frequency modulators can in general be subdivided into two classes, the reactance-tube frequency-modulated oscillator circuit, and a phase-modulation system, in which the sidebands of an amplitude-modulated signal are combined with a carrier which is shifted 90° in phase.

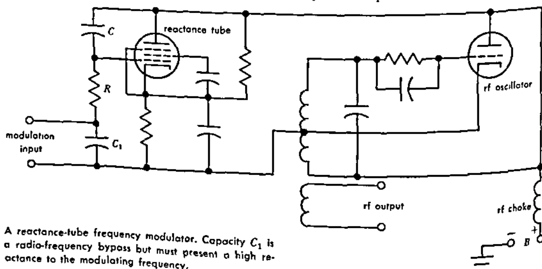
The first method is simple in principle, but auxiliary circuits must be added to control the average frequency in order to provide adequate carrier-frequency stability. The second system can originate with a crystal-controlled oscillator, and good frequency stability is obtained automatically; however, the degree of phase-modulation that can be obtained is usually relatively small This makes it necessary to employ several subsequent stages of frequency multiplication to obtain the desired frequency deviation. An improvement in the phase-modulation method is provided by a special tube called the [illegible]

[illegible] when the transmitted signal has a frequency spectrum which is equal to the bandwidth assigned for the particular channel and for which the receiver is designed The modulation index in a frequency-modulated wave is defined as the ratio of the frequency deviation of the carrier to the modulating frequency. The total bandwidth occupied by a frequency-modulated signal is approximately equal to twice the sum of the frequency deviation and the modulating frequency. In a pure frequency-modulated wave, the frequency deviation must be proportional to the peak amplitude of the modulating signal and independent of the modulating frequency. *See* FREQUENCY MODULATION; MODULATION; PHASE MODULATION.

A simple and satisfactory method for producing frequency modulation involves the use of a reactance tube to change the frequency of an oscillator. A typical circuit is shown in the illustration. A vacuum tube is connected across the frequency-determining element of an oscillator in such a way as to act as a variable reactance. The frequency-determining circuit of the Hartley oscillator is shunted by the plate-to-cathode impedance of the pentode arranged to present a variable reactance; the tube so connected is called a reactance tube because the current drawn by it is in quadrature to the voltage which appears across it. The magnitude of this current depends upon the amplification of the tube and can be varied by the modulating voltage.

The radio-frequency voltage appearing across the resistor R is 90° out of phase with respect to the plate-to-cathode voltage if the resistance R is small in comparison with the reactance of capacity C. This causes the plate current to be 90° out of phase with the plate-to-cathode voltage, and therefore the ratio of plate-to-cathode voltage to the plate current is reactive. The magnitude of this reactance can be changed by applying the modulation voltage to the control grid which will vary the current at the modulation rate. The variation of the reactance causes the oscillator frequency to change accordingly; if the changes in frequency are not large, it will be nearly proportional to the amplitude of the modulating voltage.

A reactance-tube modulated oscillator may develop some amplitude modulation because of the



A reactance-tube frequency modulator. Capacity C_1 is a radio-frequency bypass but must present a high reactance to the modulating frequency.

presence of a small resistive component unavoidably present in shunt with the variable reactance. Such amplitude modulation can be removed by passing the signal through a limiter stage or stages in which the output is independent of the input. If the output of the frequency-modulated transmitter is obtained after amplification by a number of Class-C power-amplifier stages, this limiting action will be provided automatically.

Reactance-tube modulator systems usually begin with the generation of power at a low level so that the reactance tube can operate at a correspondingly low level (in the range of a few watts). The modulator supplying the signal voltage to the reactance tube can be very simple, because no modulating power need be supplied. [E.L.G.]

Bibliography: H. S. Black, *Modulation Theory*, 1953; F. E. Terman, *Electronic and Radio Engineering*, 4th ed., 1955.

Frequency multiplier

An electronic circuit that produces an output frequency at an integral multiple of the input frequency. Frequency multipliers are used to obtain a stable high-frequency wave. For example, operation in the frequency-modulated broadcast band requires a high degree of stability at an operating frequency in the neighborhood of 100 megacycles (Mc). An oscillator operating directly at this frequency with the required stability is extremely difficult to build. Therefore, a lower frequency is generated in a crystal-controlled oscillator (where the frequency can be controlled to a small fraction of 1%), and a frequency multiplier is employed to obtain the high frequency desired. The multiple is an integer, but multiples which are the quotients of integers can be obtained by employing multiplication and division techniques. *See* FREQUENCY DIVIDER.

Harmonic generation. The plate current in a Class C amplifier is rich in harmonics of the input wave (*see* AMPLIFIER). The plate current consists of pulses occurring at the fundamental frequency of the driving signal applied to the grid. If a sharply tuned tank circuit in the plate circuit is tuned to a harmonic of the input frequency, the voltage across the tank will be sinusoidal and at the frequency of the harmonic. Practical considerations limit this technique to about the third harmonic, but by repetition of this process an output signal of a frequency nine times the reference frequency can be obtained.

Frequency multiplication can also be obtained by applying the reference voltage to a nonlinear circuit that generates the desired harmonic voltage and then amplifying the voltage from the nonlinear device in an amplifier tuned to the desired harmonic. *See* TANK CIRCUIT.

Multivibrator. The output voltage from an astable multivibrator is nearly a square wave and is therefore rich in harmonics. Frequency multiplication can be obtained by feeding the multivibrator voltage into an amplifier tuned to the desired harmonic. Practical considerations limit this technique to a fairly low-order harmonic, but the multiplication process may be repeated. Since an astable multivibrator cannot be made with the stability of a crystal-controlled oscillator, this approach is used *only* when precise control of the frequency is not important. *See* MULTIVIBRATOR. [H.F.K.]

Frequency-modulation detector

Detection of a frequency-modulated wave is carried out in two steps. First, the constant-amplitude frequency-modulated wave is converted into a wave in which the radio-frequency envelope varies in amplitude in accordance with the amplitude of the modulating signal; the modulation signal is then recovered from the amplitude-modulated wave by employing some simple form of an amplitude-modulation detector. The portion of the frequency-modulation detector circuit in which the change from a constant-amplitude wave to an amplitude-modulated wave occurs is usually called the discriminator. A number of circuits of this type are described later. *See* FREQUENCY MODULATION; MODULATION; PHASE MODULATION.

In addition to this conventional form of frequency-modulation detection, other methods are known and are sometimes used. In one of these, a frequency meter of the cycle-counting type is employed to give an indication of the instantaneous frequency. In this device, the output is proportional to the input frequency, and the variation of output with time will accurately reproduce the modulation impressed upon the carrier; it is particularly useful in applications where a linear response is required. Detectors of this type are frequently used in measurement applications where the modulation frequency is not too high. Descriptions of other known methods of detection can be found in the specialized literature.

Conversion to amplitude modulation. An amplitude-modulated wave can be produced from a frequency-modulated wave if the latter is passed through a circuit in which the amplitude characteristic varies with frequency over the range of frequencies contained within the spectrum of the frequency-modulated signal. If the amplitude characteristic is directly proportional to frequency over the range of frequencies contained in the spectrum, the resultant amplitude-modulated wave will be strictly proportional to the frequency-modulation components contained in the original wave.

The simplest circuit which approximately fulfills these requirements is a resonant circuit of either parallel or series type; the first can be operated on the side of the transmission characteristic and the latter at resonance. Although such simple circuits illustrate the principle of conversion of FM to AM, they are not usually used in practice, because the amplitude characteristic is not particularly linear, and the circuit permits spurious amplitude modulation (which may be present in the original signal) to produce unwanted output. More complicated circuits, generally employing diff

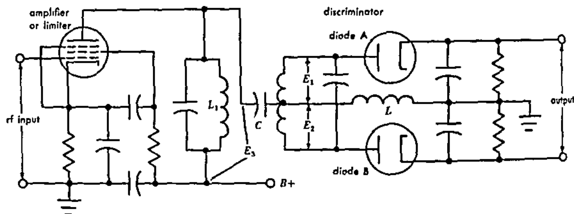


Fig. 1. Phase-shift or Foster-Seeley discriminator.

ential combination of voltages from two separate detectors, improve the linearity and help to suppress the presence of spurious AM signals.

Discriminator. The most common form of circuit used to convert FM to AM is the phase-shift discriminator, sometimes known as the Foster-Seeley discriminator. It is shown diagramatically in Fig. 1. The behavior of this discriminator can be explained as follows. Both the primary and secondary resonant circuits are tuned to the carrier frequency. The radio-frequency choke L is presumed to have a higher reactance in comparison with that of L_1, and a capacity C has a reactance which is small in comparison to the resonant impedance of the secondary. Therefore, the center of the secondary can be considered electrically connected to the top of the primary. The vector relations between the voltages of the secondary ($E_1 + E_2$) and the voltage of the primary E_3 are shown in Fig. 2; the 90° phase difference between the

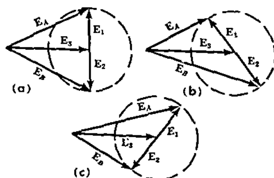


Fig. 2. Voltage relations in the phase-shift discriminator. It is assumed that the frequency is at or near the resonant frequency of the primary and secondary, making the voltages E_1, E_2, and E_3 all nearly constant in magnitude; the dashed circles show the position of vectors E_1 and E_2 as the frequency is varied. The sum of E_3 and E_1 results in the voltage E_A applied to diode A in Fig. 1. Similarly, the sum of E_3 and E_2 results in voltage E_B applied to diode B. The figures are drawn for conditions (a) at resonance, (b) above resonance, (c) below resonance. At frequencies far away from resonance, all voltages become small.

two is due to the inductive coupling between the two circuits at resonance. The phase relations between these voltages will differ with frequency. This is indicated in the diagrams for two cases, one above and another below the resonant frequency. Thus the radio-frequency voltages applied to the diodes will vary with frequency in the manner indicated.

At resonance the voltages rectified by the two diodes are equal so that the output from the detector, being the difference between the outputs of the two individual diodes, is zero. As the frequency is varied, one of the diodes receives a greater voltage than the other, producing a net difference in the output. The resultant voltage output versus frequency diagram, called the discriminator characteristic, is shown in Fig. 3. In the vicinity of resonance and over a large region between the two peaks, the output is remarkably proportional to frequency.

Practical frequency-modulation detectors of the kind described can be simplified by grounding the cathode of one of the diodes and eliminating the radio-frequency choke L. When properly designed, the phase-shift discriminator produces a nearly exact linear relation between the frequency and output over the frequency range between the two peaks, the range being determined by the selectivity factors of the resonant circuits.

A phase-shift discriminator is an excellent detector for frequency-modulated signals, provided that the carrier does not deviate greatly from the central portion of the characteristic. Because the magnitude of the output is directly proportional to the magnitude of the radio-frequency input, spurious amplitude variations of the carrier present in the input are not suppressed. When this detector is used in radio receivers, it should be preceded by amplitude limiters.

Balanced frequency detector. Another form of simple discriminator circuit is shown in Fig. 4. The radio-frequency signal is applied to the detector circuit from the primary of a radio-frequency transformer to two identical secondaries, one tuned below and another above the resonant frequency. Separate diode rectifiers demodulate the signals

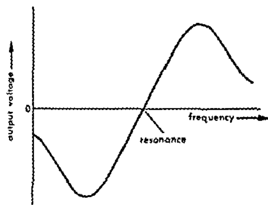


Fig. 3. Discriminator characteristic of a phase-shift discriminator, showing the relationship between the output voltage and the input frequency.

which are applied to them, and the output is obtained by taking the sum of the voltage between the two cathodes. The operation of this circuit can be understood by reference to Fig. 5, where the dotted lines show the output of each detector vs. frequency and the solid line shows the combined output.

The net result of this detection process resembles that of the phase-shift discriminator described above. Neither detector responds to pure amplitude modulation of an incoming carrier because of the balanced effect. However, the balanced detector does not suppress the spurious changes in amplitude of the applied frequency-modulated signal, nor that of the random noise which may be present. For satisfactory use in radio communication practice, it should be preceded by limiting amplifiers.

Ratio detector. A ratio detector is a modified Foster-Seeley circuit arranged to provide limiting action against changes in the amplitude of the frequency-modulated wave. It is called a ratio detector because the sum of the output of the two diodes remains essentially constant, within the response bandwidth of the resonant circuits, whereas the *modulation output derived from changes in input* frequency results from the variation in the ratio of the two diode voltages. The circuit of the ratio detector is shown in Fig. 6; (*a*) shows the circuit in its basic form for explanatory purposes, whereas (*b*) shows a modified circuit in a practical form.

The ratio detector differs from the Foster-Seeley discriminator in two respects, the polarity of one of *the detectors is reversed, and the reservoir capacitor* C_2 is added. From the viewpoint of the basic discriminator action, the reversal of the polarity is unimportant; by taking the sum of the two output voltages, as is shown in Fig. 6*a*, the shape of the

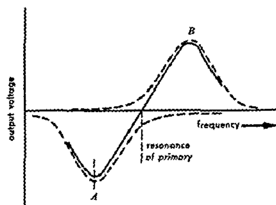


Fig. 5. Discriminator characteristic of a balanced frequency detector. Dashed curves indicate the voltages appearing across the secondaries *A* and *B* shown in Fig. 4; the solid curve, the output from the balanced detector.

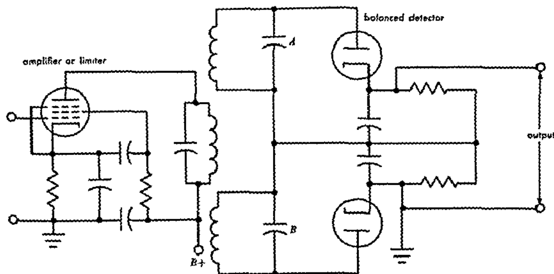


Fig. 4. Circuit diagram of a balanced frequency detector. The primary of the radio-frequency transformer is tuned to resonance and the two output secondaries are detuned symmetrically.

discriminator characteristic is found to be virtually identical to that derived from the Foster-Seeley circuit. Because the output voltage is derived by taking the two voltages in parallel, instead of in series, the ratio detector output is one-half the value produced by the Foster-Seeley discriminator. The presence of the capacity C_2 is unimportant insofar as the discriminator characteristic is concerned. This can be verified by observing that the voltage across C_2 remains practically constant as the frequency is varied in the vicinity of the resonant frequency; this can be confirmed by examining the vector diagram shown in Fig. 7.

The ratio detector can be made to exhibit self-limiting characteristics by the proper choice of the capacitor C_2 and the series resistors R_3 and R_4. If the reactance of the capacity C_2 is made negligibly small in comparison with R_2 at the lowest modulation frequency, fluctuation in the magnitude of the frequency-modulated signal can be compensated for by the proper choice of the series limiting resistors R_3 and R_4. This is due to the dynamic limiting action of the circuit; if the magnitude of [illegible] increase suddenly, a large charg- [illegible]

responding reduction in their effective Q values. A change in the Q of a circuit causes the phase relations between the secondary and primary voltages to change, the decrease of the Q causing a smaller phase shift. The vector diagrams show that a sudden increase in the signal reduces the phase shift between the voltages as mentioned and causes the output of the detectors to be reduced. By properly proportioning the circuit elements, the effect of increasing the voltage can be compensated for by the change in phase shifts, making the output

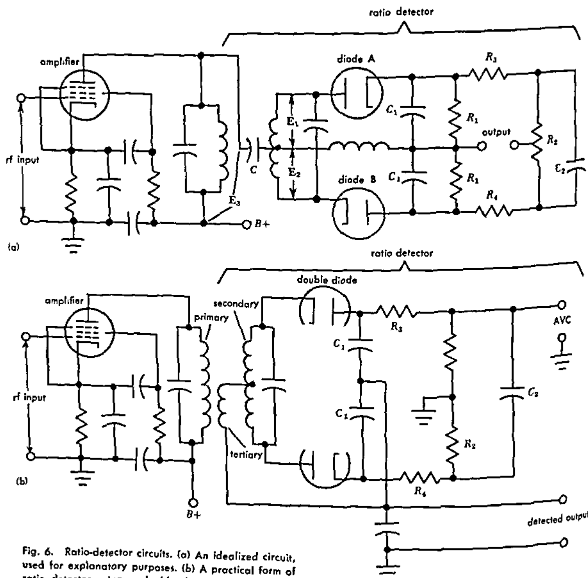


Fig. 6. Ratio-detector circuits. (a) An idealized circuit, used for explanatory purposes. (b) A practical form of ratio detector using a double diode; this circuit also provides means for deriving automatic-volume-control voltage.

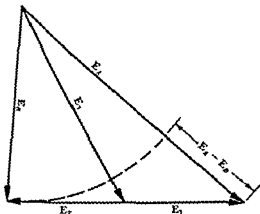


Fig. 7. Vector diagram showing the relationship of voltages in a ratio detector. Symbols refer to voltages indicated in an idealized ratio detector shown in Fig. 6a. Output of the ratio detector is obtained by taking the difference between E_4 and E_8 after each is rectified. This difference is indicated in the diagram.

insensitive to amplitude variations that exist in the signal.

A practical form of this circuit is shown in Fig. 6b which differs from Fig. 6a in a number of respects. In the form shown, the diode load resistor R_1 is eliminated because the capacitors C_1 can be made to discharge through the resistor R_2. The common radio-frequency voltage E_3 can be supplied to the two diodes by means of a tertiary winding. Automatic-volume-control voltage can be derived from one of the diodes as indicated.

The ratio detector can be made nearly identical in its linearity to the Foster-Seeley discriminator by use of a sufficiently wide discriminator characteristic. It has the advantage of good sensitivity, an input voltage of 10–100 millivolts at the grid of the driver stage being generally sufficient, whereas the Foster-Seeley discriminator usually requires an input of approximately 1 volt. The limiting action of the ratio detector is usually found somewhat superior to that which can be obtained with a Foster-Seeley discriminator and an additional separate limiter stage. The dynamic limiting provided by the ratio detector cannot suppress slow variations in signal strength, and for this reason it is necessary to employ automatic volume control when this circuit is used in radio reception. It is frequently used for the sake of simplicity, the principal advantage being that it permits the saving of a stage of amplification-limiting in the receiver.

Single-stage detector. A circuit diagram of a single-stage detector employing a heptode tube is shown in Fig. 8. This circuit is capable of providing an audio-frequency output proportional to the input frequency without the need of the usual rectifiers. When properly designed and operated at a sufficiently high input level, the circuit eliminates the residual amplitude variations almost entirely.

The operation of this detector can be explained by reference to Fig. 8. The cathode and the first two grids of the heptode are connected as a Hartley oscillator, tuned to oscillate at the carrier frequency. The oscillator is adjusted to operate as a Class-C oscillator, causing the current to flow only at the peak of each cycle, forming a series of short pulses in the output. The fundamental component of the plate current is passed through another resonant circuit also tuned to the carrier; the second resonant circuit is so adjusted and coupled to the main oscillator circuit that the impedance in the first circuit due to the second is purely reactive. The changes in the plate current vary the magnitude of this reactance and permit the frequency of the oscillator to change without affecting its amplitude.

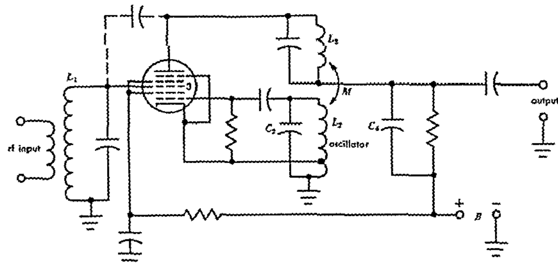


Fig. 8. Single-stage frequency-modulation detector using a heptode tube. Inductances L_2 and L_3 are coupled, introducing into L_2C_2 circuit a variable reactance.

The frequency-modulated signal to be detected is introduced into the heptode through grid 3, the oscillator circuit being shielded from the direct action of the incoming signal by the grounded screen grids surrounding grid 3. The presence of the input signal causes the anode current to change in magnitude, thus affecting the oscillation frequency. If the input signal is large enough, the oscillator changes its frequency and locks into a nearly fixed phase with respect to the incoming signal. In a locked oscillator the peak current is independent of frequency; the average current, however, is proportional to frequency because the number of pulses per second in the plate circuit is equal to the input frequency. A condenser C across the output develops the output voltage which is determined by the average current and is, therefore, proportional to the frequency of the input signal.

The only effect of the input signal is to pull the oscillator into step with the signal, that is, the oscillator becomes synchronized with the input signal, the magnitude of oscillations being independent of the input signal. If the signal becomes too weak, the oscillator will not synchronize and the detector will fail to operate.

The circuit functions as a linear frequency detector with practically no response to the residual amplitude modulation in the signal. In practice, it is important to minimize the presence of stray coupling between the input signal and the oscillator; otherwise, this can cause the detected output to be amplitude sensitive. This can be done by using specially designed heptode tubes in which the interelectrode capacity is made sufficiently small, or by using neutralizing circuits as indicated. [E.L.C.]

Bibliography: S. Goldman, *Frequency Analysis, Modulation, and Noise*, 1948; F. E. Terman, *Electronic and Radio Engineering*, 4th ed., 1955.

Frequency-modulation radio

Radio communication employing frequency modulation of a radio-frequency carrier wave to convey the desired information. In frequency modulation (FM), the frequency of the radio-frequency carrier wave is made to vary in accordance with audio or other modulation signals to be transmitted. The amplitude of the carrier wave remains constant, resulting in high efficiency of the transmitter which can operate at continuous, full-power output. *See* FREQUENCY MODULATION; RADIO.

Frequency modulation and its related phase modulation (PM) are used for most radio communication and aural broadcasting on frequencies above 30 megacycles (Mc). The main classes of radio [illegible]

sound; telemetry; and frequency-shift telegraphy.

Vehicular telephony. The versatility and economic practicality of FM or PM radio has led to its widespread use in automobiles, trucks, railway trains, buses, and harbor marine craft and others for direct telephony between other vehicles and fixed stations. Allocations for such services in the United States include frequencies between 30–50 Mc, 148–174 Mc, and 450–460 Mc. Need for these services has been so great that there has been constant effort to reduce emission bandwidths to permit more channels for more classes of users. Most government and commercial systems operate in [illegible]

numbers of associated stations to occupy one frequency. *See* TRANSMISSION THEORY AND METHODS.

Radio waves at these frequencies are sufficiently limited in their range to permit extensive geographical duplication of each channel (frequency assignment). Only in the 30–50 Mc band is there occasional long-distance interference during certain portions of the 11-year sunspot cycle when ionospheric conditions support occasional specular wave reflection over long distances.

Channel assignments in the United States are for 50-kc bandwidths, with a rapid conversion in progress (1959) to 25-kc channels in the two lowest bands listed.

Vehicular-mobile transmitters are normally below 100-watt output power, radiated from vertical whip antennas mounted on the vehicle. Most stations use from 10 to 60 watts. Fixed stations communicating with vehicles are usually of 250-watt output or less, and use more efficient antennas located at maximum feasible height above ground.

Frequency modulation is particularly useful for this service because of its constant level of signal from a receiver as the vehicle moves about. Instead, the background noise level varies inversely with carrier strength. The FM capture effect is important in suppressing weaker interfering signals, for any unwanted carrier with an amplitude less than one-half of the wanted carrier is suppressed in some degree. In the reverse case, an unwanted carrier can capture and suppress the wanted one if the amplitude ratio exceeds 2. This effect limits the utility of frequency modulation in tactical military communications.

Long-distance radio relaying. Frequency modulation is used for most radio-relay microwave systems for overland telegraph, telephone, and television transmission and for tropospheric forward scatter relay systems. Where signals must be transmitted over long distances through series of radio repeater stations, it is important that signal levels remain constant despite typical signal fading. FM systems have the property of maintaining constant signal and causing the noise levels to vary with the fading. Frequency modulation reaches its highest economic importance in this domain of its use, one little known outside engineering circles.

Radio relays are used for simultaneous transmission of up to hundreds of telephone conversations

over a trunk route. Therefore, the modulation bands are from 30 to as much as 8000 kc, divided into 4-kc portions for each voice channel. A voice channel is subdivided into several narrower bands for telegraph at standard teleprinter speeds. The radio frequencies and the frequency deviations used for modulation differ greatly. The greater the emission bandwidth, the higher the operating frequency must be to permit accurate transmission through the natural selectivity of the electrical circuits.

Fading occasionally causes a signal to fall below a critical minimum of receiver input power known as the FM improvement threshold. Below this point, all channels go into noise, even though the S/N ratio may be large just above this threshold—the point where peak carrier voltage is twice the peak value of noise at the input to the detector. For carrier voltages above this break point, FM noise-quieting takes effect and noise decreases linearly with input voltage. Radio-relay systems are engineered for specified S/N ratios in the worst channel for some high percentage of time, so the break point is seldom reached. When it occurs, it may be so brief as to be unnoticed by speakers but may cause errors in teleprinting or data transmission.

Telecommunications by tropospheric forward scatter radio relaying techniques have special need for FM techniques for maintaining stable signal levels throughout the extreme fading ranges typical of this mode of wave propagation.

FM broadcasting. This service appears in two forms—aural broadcasting in the band 88–108 Mc, and as the sound portion of television broadcasting. The English and French television systems are exceptions; they use amplitude modulation for sound.

In the United States, FM broadcast standards specify channel spacings of 200 kc, audio modulation band of 15 kc, and maximum frequency deviation of 75 kc (deviation ratio 5), making the emission bandwidth 150 kc. Actually, the modulation system is hybrid frequency-phase modulation obtained by the use of preemphasis circuits which provide pure frequency modulation for program frequencies up to 1000 cycles per sec (cps), and pure phase modulation above 5000 cps, with gradual transition between. A substantial improvement in S/N ratio results. The 15-kc modulation capability (the upper limit of acute perception) allows high-fidelity system performance when receivers of comparable quality are used. Horizontally polarized radiation is used for transmission in the United States and most of the world. See RADIO BROADCASTING.

Television sound channel. Television sound transmission differs from that for aural broadcasting in two main respects: (1) the emission frequency is associated with the video transmission in the channels assigned to television services, and (2) the frequency deviation is normally 25 kc (50-kc emission band). See TELEVISION.

FM telegraphy. Telegraphic transmission (teleprinting, binary digital data transmission for computing machine data) is largely by means of frequency-shift (FM) signaling, as distinguished from on-off (AM) signaling used since the earliest direct-current telegraph. The elements of coded characters are transmitted by transitions from a mark frequency to a space frequency, analogous to on and off switching. Both radio and wire telegraphy are trending toward complete conversion to frequency-shift. The bandwidth of emission is proportional to the rate of signaling and the amount of shift between the two states.

Telemetry. The transmission of information or data from a remote or inaccessible location (for example, a missile in flight) often is possible only by radio, and FM is used almost exclusively for the same reasons that it is preferred for other systems. Frequency modulation is usually employed for subchanneling, or multiplexing, where many different data from a multiplicity of instruments are sent in "real time," that is, as they occur. For multichannel telemetry, the total frequency deviation of the carrier wave is the statistical sum of the deviations assigned to each channel, which in turn uses frequency modulation of a subcarrier of much lower frequency, in much the same way as FM telegraph, if the data are in digital form. Analog data can be transmitted by continuous analogous changes in subchannel modulation frequency between two limits.

Telemetry enables simultaneous transmission and recording of vast numbers of instruments that are read automatically. See TELEMETERING.

[E.A.L.]

Fresh-water ecosystem

An ecosystem is the functional unit of ecology, consisting of living organisms and the nonliving environment interacting upon each other to produce an exchange of energy and materials between the living and nonliving components. A lake or pond in its entirety is such an ecosystem. A stream is also an ecosystem, although less well differentiated. Limnology is the comprehensive study of all the components of inland aquatic ecosystems and their interrelationships. See ECOLOGY; LIMNOLOGY.

The energy of fresh-water ecosystems is derived mainly from photosynthesis accomplished by the algae suspended in the water and by the higher plants and algae growing on or in the bottom. A variable proportion of the total energy available is derived from allochthonous organic matter, such as leaves and pollen, produced by terrestrial communities. The materials cycled within an ecosystem are derived ultimately from the weathering of rocks and the leaching of soil in the watershed, and to a lesser extent from the air.

Fresh-water habitats are conveniently divided into a lenitic or basin series, such as lakes, reservoirs, ponds, and bogs, and a lotic or channel series, such as rivers, streams, brooks, springs, and ground water. The lotic series is distinguished by a continual flow of water in one direction. The limnology

of the lenitic series of habitats is better understood because of the greater amount of research done on them.

LENITIC ECOSYSTEMS

The biota and biological processes of lakes and ponds operate within a framework of physical factors imposed by the external climatic environment, of which the most important are temperature stratification, light stratification, and waves and currents.

Temperature stratification. Heating of natural waters occurs mainly by the direct and rapid absorption of infrared radiation of wavelength >0.76 μ. Because 1 meter of distilled water absorbs 90% of this radiation, the direct heating of water is important only in a relatively thin surface layer. Heat at greater depths has been transported there largely by wind-generated currents and turbulence. All heat in a lake in excess of 4°C is commonly referred to as wind-distributed heat.

Two water masses of different temperatures offer a resistance to mixing by the wind which is proportional to the difference in density between them. Because of the nature of the temperature-density relationship of water, the resistance to mixing is greater at high than at low temperatures. As a consequence, in most lakes of sufficient depth, 10 meters or greater, the wind is unable to circulate the entire lake in summer. Therefore, the lake becomes stratified into an upper, warm, freely-circulating zone, the epilimnion; a lower, cold, relatively noncirculating zone, the hypolimnion; and an intermediate zone of rapid temperature change, the metalimnion, or thermocline in North American usage.

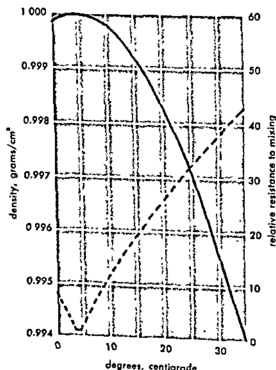


Fig. 1. Density of pure water as a function of temperature (solid line). Maximum density is at 3.94°C. The dashed line shows the relative resistance to mixing per degree difference in temperature in a water column 1 meter long. The resistance to mixing between 4° and 5° is taken as unity. (From C. D. Hodgman, ed., *Handbook of Chemistry and Physics, 37th ed.*, *Chemical Rubber Pub., 1956*)

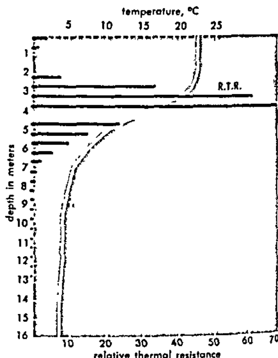


Fig. 2. A summer temperature profile (continuous curve) and relative thermal resistance to mixing (bars) for Little Round Lake, Ontario. The great thermal resistance to mixing in the metalimnion indicates a great stability. (*From J. R. Vallentyne, Principles of modern limnology, Am. Scientist, 45(3):218–244, 1957*)

The metalimnion effectively isolates the hypolimnion from surface processes, and as a result biological and biochemical processes occurring in deep water are cumulative during the period of stratification. This is one of the most important interactions of the lenitic ecosystem.

Depth of the metalimnion varies with the size and shape of the lake, the degree of protection from wind action, and the wind regime during the warming phase of the annual cycle.

In temperate lakes there is typically a spring overturn during which the lake is actively circulating from top to bottom; a period of summer stratification as described above; a period of fall overturn after the lake, by cooling, has again become homothermal, and, especially under ice cover, a period of winter stratification. The latter is often spoken of as an inverse stratification, because the warmest water is at the bottom rather than the top. A lake not covered with ice may continue to circulate throughout the winter. *See* THERMOCLINE.

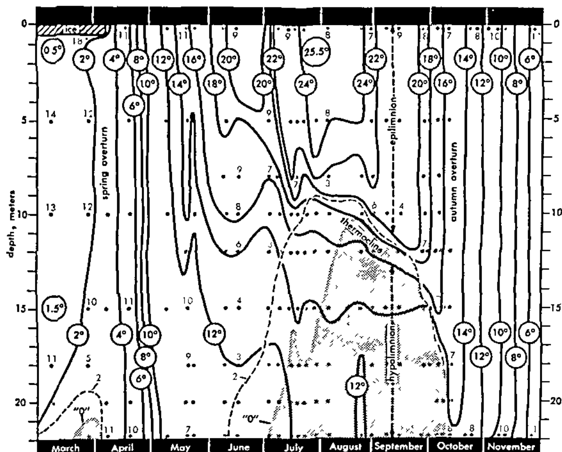


Fig. 3. Seasonal changes in water temperature and dissolved oxygen during 1906 in Lake Mendota, Wisconsin. The figures in italics are parts per million (ppm) of oxygen. The dashed line represents 2 ppm. The shaded area represents less than 0.2 ppm. (*After C. H. Mortimer from G. C. Sellery, E. A Birge, A Memoir, Univ. Wisconsin Press, 1956*)

Lake classifications. Based on these temperature relationships, an early system of lake classification was devised by F. Forel and later modified by G. Whipple. All lakes were divided into three classes, polar, temperate, and tropical. This classification depended on the annual range of surface temperature and each class was further subdivided into three orders based on the annual range in bottom temperatures. The first, second, and third orders are, respectively, the deep, moderate, and shallow lakes. One of the chief objections to the terminology employed in this system is that some of the best examples of tropical lakes are in Scotland and British Columbia. In these lakes, the surface temperature is always above 4°C.

A more realistic system proposed by G. Hutchinson and H. Löffler bases the classification on the number of periods of circulation a year, and whether the circulation occurs after warming or cooling of the lake. Dimictic lakes circulate twice a year, warm monomictic circulate once a year after a period of cooling, and cold monomictic lakes circulate once a year after a period of warming. These correspond, respectively, to the temperate, tropical, and polar lakes of the Forel-Whipple system. In addition there are the permanently frozen amictic lakes, low-altitude tropical oligomictic lakes with irregular circulation, and high-altitude tropical polymictic lakes with continuous circulation.

During turnover, most lakes circulate completely from top to bottom. Such lakes are called holomictic. In a relatively small number of lakes the water nearest the bottom has a sufficiently greater density from dissolved or possibly suspended substances so that it does not become involved in the circulation. Such lakes are called meromictic. In the noncirculating layer, or monimolimnion, of meromictic lakes the same processes occur as in the hypolimnion of a holomictic lake, but they are generally more intense and extreme because of their permanence. Sometimes the monimolimnion has a higher temperature in summer than does the overlying holomictic zone, mixolimnion, an anomalous temperature condition known as dichothermy.

Light stratification. Light is important as the source of metabolic energy in the ecosystem. The energy incorporated into organic substances by

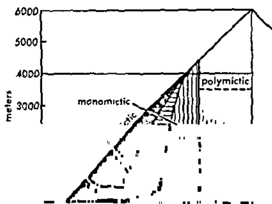


Fig. 4. The Hutchinson-Löffler system of thermal lake types. Solid black: cold monomictic. Black and white horizontal bars: transitional regions. Horizontal lines: dimictic. Crossed lines: transitional regions. Vertical lines: warm monomictic. The two equatorial types occupy the unshaded areas labelled oligomictic and polymictic, separated by a region of mixed types, mainly variants of the warm monomictic type, indicated by broken vertical lines. Horizontal scale in degrees latitude; vertical scale in meters altitude. (*From G. E. Hutchinson and H. Löffler, The thermal classification of lakes, Proc. Natl. Acad. Sci., 42(2):84–86, 1956*)

green plants through the process of photosynthesis supports the activities of all the heterotrophic organisms in the ecosystem.

Light intensity is maximal at the surface, and declines exponentially as the depth increases. Since the amount of photosynthesis accomplished is related to the intensity of light, there is a depth at which the production of organic matter by photosynthesis is equal to its utilization by respiration. This is the compensation level. The zone above is the trophogenic zone, characterized by an excess of production over consumption of organic matter. The region below is the tropholytic zone, characterized by a dominance of energy-consuming processes. In most lakes the [illegible]

the trophogenic zone.

Light passing through water also undergoes a change in its spectral composition. In any water that is sufficiently deep, the light tends to become monochromatic with increasing depth.

The factors responsible for the exponential absorption of light and its changing spectral composition are the water itself, dissolved yellow-brown color, and suspended particles or seston. The effect of any dissolved color or turbidity is to reduce the thickness of the trophogenic zone and to shift maximum transmission from the blue end of the spectrum in very pure water toward the red end of the spectrum in turbid or colored water. Dissolved color is generally more important than turbidity in producing these changes.

Waves and currents. Tides are negligible in even the largest inland lakes, being at most a few centimeters in the Great Lakes and in Lake Baikal. Wind, rather than gravity, is the important force generating waves and currents in inland waters. Such water movements erode and transport material alongshore, provide the turbulence necessary to keep microorganisms suspended, circulate the epilimnion or the entire lake, and provide for some turbulent transport across the metalimnial barrier.

The simplest waves in deep water are waves of oscillation, in which there is vertical rotation of particles without any net forward transport. The orbit of rotation decreases by one-half for each increase in depth below the surface amounting to one-ninth of the wavelength. In most inland lakes the water is essentially still at a depth of one-twelfth the wavelength.

When a wave of oscillation enters water shallower than the orbit of rotation, frictional resistance with the bottom produces an unstable configuration, and the crest of the wave now plunges shoreward as a wave of translation. It is these onshore waves that erode headlands and transport sand and other material alongshore to form spits and bars, and build subaqueous terraces and benches. *See* SHORE PROCESSES.

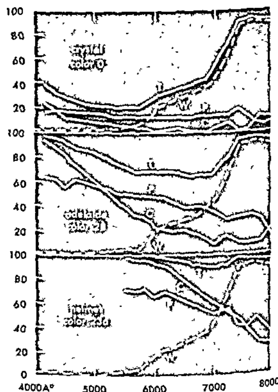


Fig. 5. Percentile absorption of light by water from three lakes in northeastern Wisconsin, ranging from visually uncolored (Crystal) to very dark brown (Helmet). Colors are listed as ppm of platinum units. *T*, total absorption; *P*, absorption due to suspended particulate matter; *C*, absorption due to dissolved color; *W*, absorption due to pure water. (*After H. James and E. Birge from G. E. Hutchinson, A Treatise on Limnology, vol. 1, Wiley, 1957*)

Fig. 6. Changes in color composition of light and intensity relative to the surface in a hypothetical lake of pure water and in four lakes of different dissolved color (in parentheses) in northeastern Wisconsin. Total intensities, relative to the incoming radiation, are represented by the areas of the circles and by the percentage values at their lower left. The shaded areas of the incoming sunlight and pure water at 1 meter represent invisible radiation, chiefly infrared. The changing position of the clock hand in each horizontal series demonstrates the selective absorption of the shorter wavelengths by dissolved color in the water. (After C. H. Mortimer from G. C. Sellery, E. A. Birge, A Memoir, Univ. Wisconsin Press, 1956)

With a steady wind from one direction the gravity-stable horizontal stratification of the lakes becomes displaced. The lake surface tilts upward slightly in a downwind direction, and the epilimnion thickens. At the same time the metalimnion tilts upward in an upwind direction, and the hypolimnion thickens here. When the generating force ceases, the now unstable density surfaces begin to rock or oscillate as standing waves. The rocking motion of the lake surface, called a seiche, dampens out quickly. The rocking movement of the thermocline, known as an internal seiche, is of greater amplitude and duration and continues long after the surface may have ceased rocking. Both surface and internal seiches can be uninodal, binodal, multinodal, or even rotational in some of the larger lakes. Such internal waves are important in affecting the vertical and horizontal distribution of organisms and in generating currents and turbulence in the otherwise relatively quiet hypolimnion. In extreme instances the hypolimnion can even reach the surface of a lake on the upwind side. *See* SEICHE.

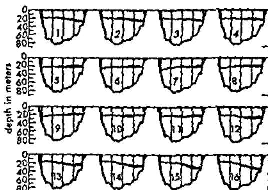


Fig. 7. Successive hourly positions of the metalimnion (the black area, bounded above and below by the 11° and 9° isotherms, respectively) on August 9, 1911, in Loch Earn, Scotland. This standing wave is an internal seiche. (After E. Wedderburn from C. H. Mortimer, The resonant response of stratified lakes to wind, Schweiz. Z. Hydrol., 15.94, 1953)

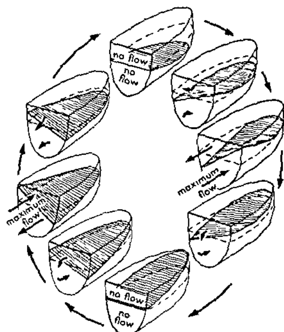


Fig. 8. Cutaway 3-dimensional model showing the effects of the earth's rotation on a *unimodal internal* seiche in a large lake, resulting in a rotational *seiche.* This model fits approximately the condition observed in Loch Ness, Scotland. The metalimnion is shaded. Arrows show the direction and relative magnitude of currents in the epilimnion and hypolimnion during various phases of the seiche. (*After C. H. Mortimer from G. E. Hutchinson, A Treatise on Limnology, vol. 1, Wiley, 1957*)

BIOTA

Within this physical framework the biota and its chemical environment react upon each other. Although all organisms in an aquatic ecosystem are mutually interdependent regardless of where they occur, they are conveniently grouped according to life habit. Microorganisms living free in the water and generally independent of the bottom are known collectively as plankton. Animals living on the bottom comprise the benthos, and the larger plants rooted in the bottom the phytobenthos. Fish and a few other large animals that swim actively in the water constitute the nekton. Sessile algae and fungi and their associated groups of microscopic animals constitute the periphyton or Aufwuchs. Organisms directly dependent on the surface film comprise the neuston. And, finally, the algae and microanimals living in the interstices of sand are known as the psammon. *See* BIOLOGICAL PRODUCTIVITY.

Plankton. This group constitutes the most characteristic assemblage in basin ecosystems and in the oceans and is the most important in terms of over-all production. Typically the plankton consists of algae, protozoa, rotifers, copepods, cladocera, and the phantom midge larva. Unlike those in the oceans, almost none of the larger invertebrates or fishes have temporary planktonic stages. Because the organisms in this assemblage have at best only limited powers of locomotion, they are kept suspended mainly by the turbulence of the water. Adaptations aiding this passive flotation are either a reduction in weight relative to the water, or an increase in relative surface area. Small size is an adaptation, and in many situations the bulk of the plankton is too small to be retained by the finest silk bolting cloth. This minute plankton is referred to as nannoplankton, in contrast to the larger net plankton.

Rate of sinking varies inversely with viscosity, which in turn varies inversely with temperature. Hence, the same organisms will sink faster in warm water than in cold. Among the planktonic cladocerans there is a summer increase in relative surface area among genetically identical generations, a phenomenon known as cyclomorphosis. The capacity for "helmet" formation is triggered by a critical temperature during development, and the helmet is maintained largely in response to turbulence in the water. Some rotifers and dinoflagellates also show increased surface area in summer.

Although a list of plankton organisms in a particular ecosystem often numbers several hundred species or more, chiefly algae, the number of species present at any one time is generally small. Thus, at any instant the majority of small to medium-sized lakes have only 1–3 species of copepods, 2–4 cladocerans, and 3–7 rotifers. In each of these

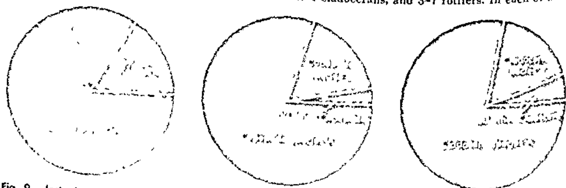


Fig. 9. Instantaneous percentage species composition of the three dominant groups of zooplankters in a typical small to medium-sized lake. (*From R. W. Pennak, Species composition of limnetic zooplankton communities, Limnol. and Oceanog., 2(3):222–232, 1957*)

Fig. 10. Seasonal succession of dominant organisms in a small permanent pond in Pennsylvania (From S. S. Bamforth, Ecological studies on the planktonic protozoa of a small artificial pond, Limnol and Oceanog., 3(4):398–412, 1958)

groups of animals the dominant species comprises about three-fourths of the total population.

Variability. A chief characteristic of plankton is its variability both from season to season and from year to year. Species succeed one another often in bewildering array. Succession seems to be governed by the ever-changing combination of temperature and light, nutrient depletion, such as silica for diatoms, and by metabolites, long-chain fatty acids, carbohydrates, antibiotics, vitamins, and other substances given off to the environment that can be self-limiting to the producing species, and antagonistic or synergistic to another species. Parasitic fungi also are claimed to have a significant effect.

A peak development of a single species is known as a pulse, and when a pulse is great enough to be apparent to the unaided eye it is called a bloom. All seasons of the year exhibit pulses of at least some species, although these are often too small to affect materially the total population. Considering plankton as a whole there is generally a peak population or bloom in spring, mainly diatoms, a reduction during the summer, a second although lesser peak in fall roughly coinciding with the autumnal overturn, followed by a decline during the winter to the lowest population of the year

Vertical distribution. The vertical distribution of plankton within a lake is the result of many physical, chemical, and biotic factors. Because of the dominating influence of light, however, most species tend to have their maximum population density within the uppermost 10 meters. Some of the larger zooplankters migrate toward the surface at night, and back into deeper water toward morning. Such diel vertical migration is particularly well developed in the deep-scattering layer of the oceans

Plankton rain. In spite of turbulence there is a continual settling of senescent and dead plankton to the bottom of the lake. This plankton rain sup-

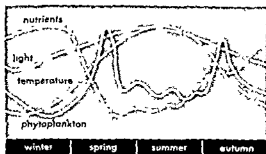


Fig. 11. Typical standing-crop curve for phytoplankton in response to temperature, light, and nutrients. Decline of the spring bloom is at least in part a response to nutrient depletion, and establishment of the autumn bloom is at least in part a response to greater nutrient availability. (From E. P Odum and H. T. Odum, Fundamentals of Ecology, 2d ed., Saunders, 1959)

plies energy to the deep-water benthos. It also serves as the chief cause of the extensive chemical differences between surface and bottom water that develop during summer or temporary stratification and under the ice in winter. The surface water tends to become depleted of nutrients, and the bottom water becomes enriched in them. Moreover, the utilization of plankton rain by bacteria and animals results in a reduction of the oxygen supply in deep water. The extent of oxygen reduction is also influenced by the amount of allochthonous organic matter that gets into deep water.

Plankton reproduction. Cladocerans and rotifers reproduce by parthenogenesis during most of the summer. This permits the rapid utilization of developing plankton pulses. During the deterioration of environmental conditions, such as the approach of winter, most of the cladocerans and many of the rotifers produce resistant bisexual or resting eggs that enable the species to endure during the unfavorable conditions. Resting eggs also facilitate the distribution of species by wind and birds. As a result, many fresh-water plankters, including species of algae, are virtually cosmopolitan in distribution. Copepods, on the other hand, reproduce almost exclusively by fertilized eggs. Some species form resistant eggs, but more commonly they survive unfavorable conditions by the encystment of immature stages. One species in Douglas Lake, Michigan, produced more than 1,000,000 cysts per square meter of bottom in the hypolimnion. One species of *Daphnia* in the Arctic produces resting eggs by parthenogenesis. *See* PARTHENOGENESIS.

Zooplankton. The majority of the zooplankton are herbivorous, and feed on bacteria, small algae, and organic detritus which they strain from the water by a variety of means Their grazing activity can effectively reduce the standing crops of phytoplankton. A few cladocerans and rotifers, the phantom midge larva, and adult cyclopoid copepods are predacious. When abundant, the zooplankton can markedly reduce the oxygen content of restricted strata.

Phytobenthos. The rooted aquatic plants form the other major photosynthetic element in fresh water. The relative importance of the phytobenthos and phytoplankton in over-all [illegible]

[illegible] benthos assumes a progressively more important role. Because of its longer life cycle it ties up nutrients for longer periods than does the phytoplankton, and it contributes more organic matter to the lake bottom and causes a faster rate of filling.

The phytobenthos typically occurs as three concentric zones around a lake. At the shoreline and in shallow water is the zone of emergent plants, followed by the zone of floating leaf plants, in turn followed by the zone of submersed plants. At still greater depths sometimes occur meadows of the moss *Fontinalis*, which is able to thrive in weaker light than that which the higher plants require. In many situations, particularly in hard-water lakes, the coarse branched algae *Chara* and *Nitella* form dense beds.

The phytobenthos and the thermal stratification are used to define the zonation of the bottom. From the shoreline to the lower limit of rooted aquatics is the littoral zone, the lower limit of which corresponds roughly to the compensation level. Between the littoral zone and the top of the metalimnion is the sublittoral zone, and at still greater depths is the profundal zone.

Periphyton. One of the chief functions of the phytobenthos in the over-all economy of an ecosystem is to increase the amount of colonizable substrate in the trophogenic zone. The substrate in the littoral zone is often thickly overgrown with a mat of sessile algae and fungi, and a large number of associated and dependent animals such as protozoa, *Hydra*, microcrustacea, rotifers, oligochaetes, insect larvae, and snails. This community of organisms, complete with producing and consuming elements, is almost the same as an ecosystem within an ecosystem. It helps contribute to the high productivity and the diversity of microhabitats of littoral regions.

Benthos. The great differences between the littoral and profundal benthos are controlled in part by the environmental conditions and in part by the life cycles of the individual species. In the littoral zone with its coarser sediments, wave action, a continuous supply of oxygen, warm temperatures, plus light and plants, diversity of species is the chief characteristic. Insects and mollusks are dominant, many of them confined to the littoral. All other groups of fresh-water animals also occur here. In the profundal zone, on the other hand, soft sediments, virtual absence of light, low temperatures, and potentially severe chemical conditions, especially reduction in oxygen content, severely restrict the species composition. Midge larvae, oligochaetes, and fingernail clams are the most characteristic of the larger animals. In the ooze-film assemblage there may be a great diversity of microscopic animals.

The profundal benthos is largely dependent for its livelihood on the plankton rain from above. Most of the organisms feed either directly on this detritus or on the bacteria that are decomposing it. Largely confined by steep chemical gradients to the uppermost few centimeters of the sediment, these organisms can thoroughly work through the sediments in much the same way that earthworms do through the soil. The relatively few species are sometimes present in tremendous numbers of individuals, and the total population density often exceeds that of the littoral zone.

The general absence of sessile animals in the littoral zone is one of the chief biotic differences between the oceans and inland waters. Seasonal redistribution [illegible] commonly occurs. Insect [illegible] or pupa- [illegible] oreward

to breed. Emergence of insects in spring and summer helps to produce a minimum littoral population at this time. The maximum occurs in the winter.

Psammon. In the interstices of sand grains along lake shores is a highly specialized community consisting chiefly of pennale diatoms, bacteria, protozoa, rotifers, copepods, and primitive copepodlike crustaceans of the order Mystacocarida. Even though the sand is shifting under the influence of waves and currents, the psammon organisms are quite distinct from those in the overlying water. The shallow penetration of light into the sand enables the diatoms and other algae to make the community largely self-sustaining. *See* MYSTACOCARIDA.

Nekton. In temperate waters fish dominate the nekton. Species requiring cold water and high levels of oxygen such as trout, coregonids, and burbot can survive in the hypolimnion of many of the larger lakes. Species tolerant of or requiring higher temperatures occur in the surface waters. Among these, certain species live offshore and others occur in the littoral region, either among weed beds or among rocks of shores exposed to strong wave action. Nearly all species, however, regardless of where the adults occur, spawn in shallow water. Here the young find an abundant food supply, and cover to help protect them from predators. In tropical lakes decapod shrimps of the families Atyidae and Palaemonidae can be important components of the nekton.

Higher vertebrates. Frogs, salamanders, turtles, snakes, crocodilians, various aquatic birds such as loons, herons, kingfishers, mergansers, and ospreys, and various mammals such as muskrat, beaver, racoon, mink, and otter are peripheral members of the aquatic biota. Many of them are essentially terrestrial animals that derive only a part of their sustenance from the water, and because they feed mainly on fish, their total demand on the energy resources of the ecosystem is relatively minor. For these reasons they are often omitted when aquatic ecosystems are considered.

Neuston. A minor life-habit assemblage is that associated with the surface film. Here belong the various insects that run, skate, or swim on the surface of the water, and some truly aquatic organisms such as bacteria, fungi, and certain algae that are suspended from the underside of the film. Quite a number of aquatic animals such as snails, *Hydra*, aquatic bugs, and beetles regularly or adventitiously occur on the underside of the film for brief periods, and some such as anopheline mosquito larvae feed on the microorganisms attached there.

Trophic relationships. The phytoplankton, phytobenthos, and algae of the periphyton and psammon are the producers. Their raw materials are light energy, carbon dioxide, and soluble inorganic sources of nitrogen, phosphorus, and sulfur. In the process of photosynthesis they manufacture organic compounds and give off oxygen to the environment.

The animals are the consumers. They utilize the organic compounds elaborated by the plants and modify them for their own metabolic needs. Through their activities they are continually converting chemical energy to heat energy which is irretrievably lost to the environment. Because energy can enter the system only at the producer level, there is a continuous and steady decrease in supply of bound energy with each successive level of utilization. In their metabolism animals use oxygen and give off carbon dioxide.

Bacteria are the reducers or transformers. They ultimately mineralize the metabolic wastes and dead organisms to liberate the nutrients in a form which may be used again by the producers. Thus, in an ecosystem there is a continuous recycling of materials. Aerobic bacteria utilize free oxygen and [illegible]

CHEMISTRY

Changes in the chemistry of the water reveal much about the metabolism of the ecosystem and about general limnological relationships.

Oxygen. Dissolved oxygen, more than any other single substance, has advanced knowledge of limnology by permitting diagnosis of what is happening in an ecosystem.

The epilimnion, as a result of circulation and photosynthesis, is generally saturated with oxygen. In the hypolimnion, however, decomposition of the plankton rain and other organic matter results in a utilization of oxygen, which cannot be replenished until the next overturn. The extent of this depletion depends on the amount of organic matter being furnished to the hypolimnion, that is, the rate of production in the epilimnion, the volume of oxygen available for utilization, which is a morphometric factor roughly equivalent to the volume of the hypolimnion, and the temperature of the hypolimnion, a climate dependent factor, as influencing the rate of metabolism.

Classification based on productivity. Based on the rate of production in the epilimnion, lakes have been classified as oligotrophic or unproductive and eutrophic or productive, with a continuous spectrum between. Off to one side are the brown-water or dystrophic lakes. A. Thienemann defined an oligotrophic lake as one which has more oxygen in the hypolimnion than in the epilimnion, and a eutrophic lake as the reverse. This, however, equates the level of production with the hypolimnetic oxygen supply and ignores the other two controlling factors B. Aberg and W. Rodhe have resolved this situation by proposing descriptive terms for oxygen distribution without reference to level of production. An orthograde oxygen curve exhibits little or no decline in the hypolimnion. A clinograde curve shows a marked decrease, having a general shape similar to that of a temperature curve. Heterograde curves are special types, having either a maximum, plus heterograde, or a minimum, minus heterograde, oxygen content in the metalimnion. Orthograde and

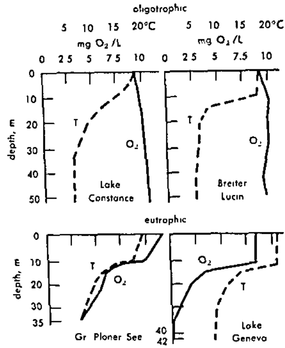


Fig. 12. Oxygen and temperature stratification in one Wisconsin and three European lakes. The two upper oxygen curves are orthograde, and the two lower ones clinograde. (*From F. Ruttner, Fundamentals of Limnology, Univ. Toronto Press, 1953*)

clinograde oxygen distributions correspond, respectively, to oligotrophy and eutrophy as the terms are still commonly used.

Change in productivity. Whatever the causes, most lakes, with time, experience a decline in the oxygen content of the hypolimnion. This is commonly the result of increasing production in the epilimnion and of the gradual accumulation of sediments, which brings about a reduction in the volume of the hypolimnion. The rate of change can be accelerated by man through pollutional and agricultural enrichment, resulting in a higher rate of production in the epilimnion.

As the oxygen content declin[illegible]

mer may have as many as 200 species of profundal benthos exclusive of protozoa, whereas with reduction in oxygen content the number of species is reduced, until in the most severe cases there are no higher animals at all except perhaps the facultatively anaerobic phantom midge larva. Thienemann first demonstrated this relationship for midges in the lakes of northern Germany and pointed out that lakes with only slightly reduced oxygen content have *Tanytarsus*, those with greatly reduced oxygen *Chironomous*, and those with intermediate levels *Stictochironomus* or some other genus. *Tanytarsus* has been used as an indicator of oligotrophy and *Chironomus* of eutrophy.

With respect to fish, lakes with adequate oxygen in deep water can be called two-story lakes, meaning that there is an assemblage of cold-water fishes in deep water distinct from the warm-water fishes in the epilimnion. Reduction in oxygen ultimately eliminates the cold-water fishes to produce a one-story lake. In North America the cisco has been able to persist in some lakes with a plus-heterograde oxygen distribution but with no oxygen in the hypolimnion.

A critical point in lake ontogeny is the disappearance of oxygen from water in contact with the bottom. Aerobic processes are replaced by anaerobic. Methane and hydrogen sulfide are formed in quantities and get into the water. Lowering of the oxidation-reduction potential results in the conversion of previously insoluble iron and manganese to a soluble condition, allowing them to diffuse into the water. Any excess iron can create a trap for phosphate under these conditions and prevent it from returning to the epilimnion at the next overturn. Methane bubbling through the hypolimnion helps reduce the oxygen content and induces turbulence. This is known as methane circulation.

Carbon dioxide. Carbon dioxide occurs in water as the dissolved gas, as carbonic acid, and as the [illegible] of calcium and mag-[illegible]

carbon dioxide complex is the most important single control of pH in inland waters.

Unlike the oceans, inland waters do not have a constant percentage composition. However, in most waters the world over the dominant ions, calcium, magnesium, sodium, potassium, bicarbonate, sulfate, and chloride, occur in the same order of abundance: $Ca^{++} > Mg^{++} > Na^{+} > K^{+}$, and $HCO_3^{-} > SO_4^{--} > Cl^{-}$. As a result, the pH of most waters varies between 6.5 and 8.5. Higher pH occurs regularly in waters of arid regions and temporarily in hard-water lakes during vigorous photosynthesis. Lower pH is produced by humolimnic acids in dystrophic lakes and by inorganic acids, chiefly sulfuric acid, H_2SO_4, in special situations.

During photosynthesis carbon dioxide (CO_2) is withdrawn by the plants, causing the CO_2 equilibrium to shift with the production of the relatively insoluble calcium carbonate, which settles to the bottom. In littoral regions the precipitate tends to accumulate as marl. In the profundal region much or all of the carbonate can be reconverted to soluble bicarbonate by the excess CO_2 of the hypolimnetic metabolism.

A soft-water lake is one which has only small amounts of calcium and magnesium. Being dependent largely on the atmosphere for an additional supply of CO_2, such lakes are said to have a chronic CO_2 deficiency. Hard-water lakes have a CO_2 reserve in the calcium and magnesium bicarbonates, thus permitting a greater amount of production to be accomplished. Even they, however, can have all their reserve CO_2 utilized, resulting in an acute CO_2 deficiency. In such lakes some of the calcium carbonate can continue to be present in nonsettling colloidal form. This causes the methyl orange ti-

tration to give erroneous results as to the CO_2 reserve.

Other elements. Phosphorus and nitrogen are the two elements most commonly in short supply. Of these, phosphorus is generally more critical because of the nature of its supply. Nitrogen is fixed in the atmosphere during electrical storms, and in the aquatic ecosystem by certain blue-green algae and bacteria. Although these two elements and other nutrients a priori must exert control on production and succession in the ecosystem, the specific effects have usually been difficult to demonstrate in the field.

Production and productivity. The organisms present at any one time constitute the standing crop, which is the result of production but is not indicative of the rate of production. The quantity of plankton beneath a unit area of lake surface is approximately the same for oligotrophic and eutrophic lakes, although the quantity per unit volume of the trophogenic zone is greater in eutrophic lakes. Here the plankton is so dense that it limits its own production by reducing light transmission.

Measurement of productivity. Production in planktonic diatoms has been measured by collecting and counting the cells that settle into jars suspended in the water. In several lakes greatly different in size a reproduction rate of about 10% per day was necessary to maintain the standing crop at a particular level. Thus, a low population density may be the result of a rapid rate of sinking rather than a low rate of production.

Rates of production of phytoplankton are commonly estimated by measuring chemical changes developing between clear and opaque bottles containing portions of the same phytoplankton population. In the clear bottle photosynthesis and respiration occur, in the opaque bottle only respiration. The difference between the bottles represents gross production, and the difference between the light bottle and the initial value represents net production. The latter represents the organic matter and energy available for utilization by the heterotrophic components of the ecosystem. Changes measured in the bottles are oxygen, pH, electrical conductivity, and radiocarbon (C^{14}) uptake. The latter apparently measures net production directly, although a correction must be made for fixation of carbon in the dark.

Photosynthetic capacity. Since photosynthesis is dependent on light intensity, on bright days the trophogenic zone is thicker than on dull days, and it is thicker in summer than in winter. Because of the photoreduction of chlorophyll by bright light, maximum photosynthesis may be some distance below the surface of the water. From the photosynthetic capacity of the phytoplankton community present, it is possible to approximate the amount of photosynthesis accomplished over a 24-hour period by integrating the various subsurface light-intensity curves. In epilimnetic communities there is also a relatively constant relationship between chlorophyll content and photosynthesis. Because of

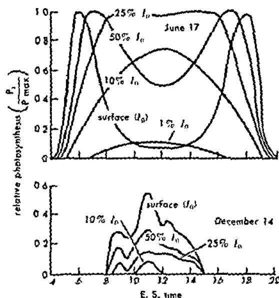


Fig. 13. Daily relative photosynthesis at various depths in the water corresponding to stated percentages of surface light intensity for typical days in summer and winter at Newport, Rhode Island. (From J. H. Ryther, The measurement of primary production, Limnol. and Oceanog., 1:72–84, 1956)

this direct dependence of photosynthesis on light intensity it is not surprising that the curve for production accomplished closely follows the curve for total illumination during the year. Maximum production occurs in summer when the standing crop is low relative to spring and fall. The biomass is kept at low levels by the grazing activities of herbivores in the plankton.

Consumer level. At the consumer level, production is more difficult to study because of the increased number of pathways into which the energy can be channeled. Under these conditions the concept of production is somewhat meaningless without reference to a particular product. In general, production is greater, although the standing crop is less, when the product organism is being utilized by some higher-level consumer. In two ponds studied, fish reduced the standing crop of their food benthos but increased its production. Total production was about 17 times the standing crop, which compares favorably with production coefficients determined for diatoms.

Much of the organic matter fixed in the epilimnion is consumed and mineralized there to enable the nutrients to be reutilized immediately. Only a relatively small proportion of the total primary production, about one-fourth in north German lakes, gets into the hypolimnion as plankton rain. However, the chemical changes induced in the hypolimnion can be used as an index of the relative levels of production in lakes. If the changes are related to unit surface area of the hypolimnion, through which the plankton sinks, the results are largely independent of the relative size of the epi-

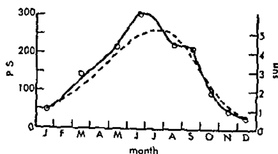


Fig. 14. Annual course of photosynthesis (solid line) in relation to solar radiation (broken line) under completely natural conditions in western Lake Erie. Note that maximum photosynthesis occurs in summer when the standing crop is not maximum (see Fig. 11). (From J. Verduin, Primary production in lakes, Limnol. and Oceanog., 1:85–91, 1956)

limnion. The two quantities most commonly measured are the rate of generation of the hypolimnetic oxygen deficit and the rate of generation of the hypolimnetic carbon dioxide accumulation. In strongly eutrophic lakes the latter measure is better, since CO_2 continues to be produced, although at a lesser rate, under anaerobic conditions after the oxygen may have long since disappeared.

LOTIC ECOSYSTEMS

A river system consists of a treelike arrangement

... in deltas and alluvial fans are morphologically analogous to a reduced root system. In limestone regions, particularly, various headwater streams can be in direct contact with an ...

... snow and ice, and from ground water. Surface drainage contributes the bulk of inorganic ...

... increase, and dissolved ...

... hydrograph. The ascending limb ...

... is dual in nature, consisting of a stream of water and all its contained substances, and a stream of heavy materials constituting the bed load, which moves along the bottom intermittently or continuously in the nature of a giant file. The molar action of this material continually erodes the stream channel deeper, and when severe, such as during a flood, can virtually eliminate the biota from large areas. Production of organisms is directly related to stability of bottom. By eroding, streams constantly tend to cut down their channels to base level.

Gradients. Except for light, gradients in lotic ecosystems are longitudinal rather than vertical. Most commonly the slope profile of a stream is concave upward, with the steepest slopes in the headwaters. This creates gradients of current velocity, coarseness of bottom, and other conditions. Because the headwaters are generally at considerably greater elevations and in more direct contact with the ground-water system, there tend to be gradients in temperature such as a warming downstream in summer and in magnitude of annual temperature variation which is least upstream.

A stream is a continuum only in a physical sense. The various longitudinal gradients interacting bring about a longitudinal succession of species and of communities. Generally the number of species in a taxonomic group increases in a downstream direction. This is known to be true of fishes in the Rhine and Colorado rivers, and of protozoa and insects in Pennsylvania. The replacement of one species of turbellarian by another in response to a changing temperature regime was first found in ...

commonly divided into four biological zones on the basis of the dominant fish species. Uppermost is a

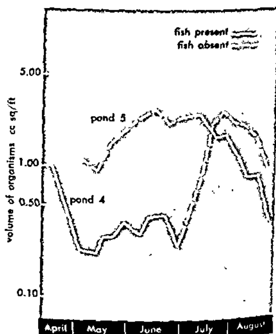


Fig. 15. Fluctuations in volume of standing crop of those benthic organisms eaten by fish in two ponds in Michigan. The action of the fish in each case was to reduce the size of the standing crop but to increase its rate of production. (From D. W. Hayne and R. C. Ball, Benthic productivity as influenced by fish production, Limnol. and Oceanog., 1:162–175, 1956)

trout zone, with steep gradient, rapid flow, close alternation of pools and riffles, and low temperatures. Below this is the grayling zone, with reduced current velocity, greater annual fluctuations in temperature, and a more open valley with the stream beginning to meander in a small flood plain. The barbel zone is in some respects transitional between the two cold-water zones above and the warm-water bream zone below, dominated by cyprinids, percids, pike, and other warm-water fishes. The several zones are associated with various stages in the physiographic cycle, from the youthful V-shaped valleys and steep gradients in the headwaters to mature, relatively flat valleys and extensive flood plains downstream. By the erosional activity of a stream there is a gradual headward migration of stream zones or habitats and their associated biotas.

Plankton. Various types of plankton occur in both lotic and lenitic environments.

Periphyton. In headwater streams the dominant producing element is the sessile algae on the bottom, which is constantly being broken loose by the current and drifted downstream. The microscopic material (plankton) in the overlying water is sometimes described as a "pale image" of the periphyton community.

Potamoplankton. In more sluggish sections of a stream a true plankton can develop. Most of the organisms present are the same as those occurring in lakes and ponds. Their source is backwaters, sloughs, and flood-plain pools. Plankton leaving a lake tends to be rapidly removed from the stream by entrapment in the water immediately in contact with the bottom or other substrate. The plankton of the Rhine, however, is dominated for long distances by organisms derived from the Swiss lakes via the Aare River.

Benthos. The benthos of still-water sections is similar to that of lakes. In currents the two distinctive communities are the stone fauna and the moss fauna. Organisms in the stone fauna tend to be greatly flattened for greater contact with the substrate. Many are directly attached to the substrate, or they live in webs and cases that are attached. In temperate latitudes the net-winged midges and in the tropics fishes and tadpoles have suckers for attachment. Those organisms when wrested loose and drifted by the current are known as syrton. They constitute another example of the downstream transport of organic matter, and they are important in colonizing new or depopulated areas. The tendency to be swept downstream is compensated for in part by a positive rheotaxis, which directs the animals into the current and makes them move upstream against it. Animals of the moss fauna are small and well provided with grappling structures. The moss serves primarily as a substrate, with relatively few organisms feeding on it directly.

Phytobenthos. Rooted aquatics occur abundantly in areas of reduced current. Their stems and roots further reduce the current velocity and accumulate fine sediment, which changes the nature of the stream bottom and helps delay the downstream transport of nutrients.

Nekton. Fishes which require low-temperature and high-oxygen environments tend to occur in the headwaters, and the more eurytropic fishes occur in the lowland sections. Fishes in swift currents tend to be almost round in cross section, whereas many of those in quiet water are strongly compressed from side to side.

Many, perhaps most, stream fishes have a restricted home range in which they spend their entire lives. If displaced downstream, for instance during a flood, they are able to find their way back. In the long-eared sunfish this is accomplished by the sense of smell more than by sight.

Continental fish faunas are dominated by cypriniform fishes, consisting of the carplike fishes such as the characinoids, gymnotoids, and cyprinoids, and catfishlike fishes. Approximately 80% of the fishes of southeast Asia are of this order, 50% of the Great Lakes fishes, and more than 80% in the Amazon. Peripheral fish faunas have smaller percentages of cypriniforms.

Production relationships. R. Butcher attempted to estimate productivity by the number of algal cells settling on glass slides submerged for 28 days. Oligotrophic streams yielded fewer than 2000 cells/mm^2, whereas eutrophic streams yielded 2500–10,000 cells/mm^2. Regions with moderate organic pollution yielded higher numbers, up to 100,000/mm^2. J. Yount has found for sessile diatoms that as the population density increases, the number of species represented decreases.

In the upper sections of streams, riffles are generally more productive than pools. Current produces a eutrophication effect, and animals living in a current have a higher rate of metabolism than those in still water. In mayflies the area of gill surface decreases as the current velocity increases.

Estimates of production have been made from the 24-hour cycles of dissolved oxygen and carbon dioxide with suitable corrections for diffusion. Attempts have also been made to approximate primary production from the quantity of chlorophyll per unit area of bottom. In the outflow of a spring, production and the pathways of energy through the ecosystem have been measured by a modification of the light bottle, dark bottle principle. Because the substrate is so important in primary production in streams, the upstream shallow portions may be considered a trophogenic zone which exports its surplus production, such as plankton and syrton, to a downstream, deeper tropholytic zone. Production of detritus-feeding benthonts such as midges, mayflies, oligochaetes, and small clams in the lowermost portions of large rivers is at times tremendous.

Pollutional relationships. Quantities of organic matter entering a stream from domestic sewage, a paper mill, or a canning factory, can markedly change the balance of biological processes. Bacteria attack the organic matter, resulting in a utilization of oxygen and a release of plant nutrients.

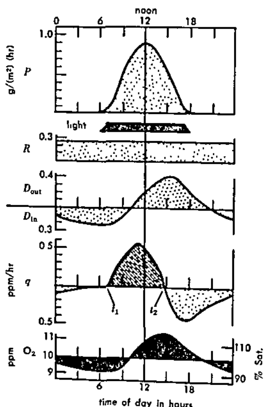


Fig. 16. Component processes in the oxygen metabolism of a section of a hypothetical stream during the course of a cloudless day. The rates for production *P*, respiration *R*, and diffusion *D* are combined in *q*. The shaded area delimited by t_1 and t_2 multiplied by the volume of flow in m³/hr and divided by the area in m² of the stream section studied gives gross primary production per day. The method has been successfully used in streams, coral reefs, and other marine littoral areas for measuring primary production. *(From H. T. Odum, Primary production measurements in eleven Florida springs and a marine turtle-grass community, Limnol. and Oceanog., 1(2):85–97, 1956)*

The extent of oxygen utilization depends on the relative dilution of the organic matter by the river water. Engineers refer to the oxygen sag curve and the process of reaeration downstream. In severe cases the oxygen can be completely utilized, resulting in a septic zone or polysaprobic zone. This is a zone of bacteria, with anaerobic processes predominating. Farther downstream conditions in the stream begin to recover, forming a zone of partial recovery, the mesosaprobic zone, and still farther downstream is the zone of complete recovery or oligosaprobic zone. In this self-purification process there is a longitudinal succession of bacteria, protozoa, algae, and insects. There is also a definite and predictable longitudinal succession of chemical conditions.

Other types of pollution such as toxic wastes and inert materials may affect the biota of a stream without changing the oxygen content. A stream is considered healthy if a great variety of microhabitats in it permit establishment of a diversity of species and trophic relationships in balance with one another. Pollution of any kind eliminates various of these microhabitats and their adapted species and often results in the increased abundance of some of those remaining. The severity of pollution has been measured by comparing the number of species in several taxonomic groups with the average numbers of species present at nonpolluted sites in the same watershed. Fishes, insects, and crustaceans are the most sensitive to pollution. Blue-green algae, bdelloid rotifers, oligochaetes, leeches, and pulmonate snails are least sensitive.

Since the early work of R. Kolkwitz and M. Marsson, attempts have been made to find indicator species for different severities or kinds of pollution. Most species, however, are sufficiently adaptable in their requirements that they can occur in more than one zone. The community of organisms present is more important than the individual species in indicating community metabolism. *See* ECOSYSTEM; ENVIRONMENT; MARINE ECOSYSTEM.

[D.G.FR.]

Bibliography: G. E. Hutchinson, *A Treatise on Limnology*, vol. 1, 1957; F. Ruttner, *Fundamentals of Limnology*, 1953; J. R. Vallentyne, Principles of modern limnology, *Am. Scientist*, 45(3):218–244, 1957; P. S. Welch, *Limnology*, 2d ed., 1952.

Freudianism

The psychoanalytic school of psychiatry founded by Sigmund Freud (1856–1939).

Premises. The school is based on the following premises: (1) the conception that sexual impulses and their transformations are the most fundamental activators of human behavior (in later versions, hostile impulses were posited as an additional second set of basic activators); (2) the doctrine of the economy of mental functioning that assumes that all behavior is an expression of a compromise between various drives and defenses that are operative, and that this economy involves an exchange and transformation of energy (libido) in a closed system; (3) the assumption that there is a tripartite mental structure, involving three conflicting institutions—an id that is the source of instinctual impulses (instincts) that are unconscious to the person, a superego that represents conscience and the sense of social prohibitions, and an ego that serves as regulator and executor between the demands of superego and id by means of integrative and synthetic functions; (4) that much of mental life stems from sources that are unconscious to the person, notably the id and portions of the ego and superego, which produce behavior that is seemingly irrational in the sense that the person cannot account for it; and (5) a doctrine of complete psychic determinism that holds behavior not to be random or accidental but rather to be produced by the interaction of psychic forces, principally unconscious.

Theory. Psychoanalytic theory is thoroughly developmental in viewpoint. Three key phases of psy-

[illegible]

[illegible] as well as for the identification with or internalization of the standards of the parent.

Psychoanalytic. Psychoanalysis tends to see the impulse life of the individual in basic conflict with the demands of society. The impulse life demands wish fulfillment; society demands restraint and adherence to its standards. There is conflict until the individual internalizes society's demands and learns how to modify his impulses accordingly. It is the developing ego that makes possible the resolution of the seemingly unresolvable internal impasse, for the ego finds ways of satisfying both sets of demands. Behavior and fantasy and dream then become joint products of conflicting demands; the object of psychoanalysis as therapy is to give the person insight into the multiple determination of his behavior. See NEUROSIS.

The requirement of compromise, the so-called internal economy of personality, leads to elaborate

[illegible]

Fig. [illegible]

[illegible] one can calculate [illegible]

$$[illegible]$$

Figure [illegible] shows the forces acting on an object with a weight of [illegible] being pushed at a constant speed to the right. Using one of the conditions for static equilibrium, the summation of forces in any direction must equal zero; the normal force is seen to be equal to the weight, and the horizontal force F required to push the object is equal to the friction force. Thus

$$[illegible]$$

The angle [illegible] is called the friction angle and it [illegible]

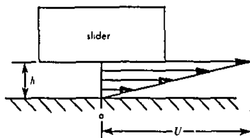


Fig. 2. Action of oil film in reducing friction.

found from

$$\phi = \tan^{-1}\frac{F_f}{F_n} = \tan^{-1}\mu$$

The viscous friction most often encountered in machines is that associated with journal or sleeve bearings In this case, the friction force acts at a radius and becomes a friction torque.

In Fig. 2 the slider is separated from the fixed body by an oil film with a viscosity μ and thickness h and is moving to the right with speed U. The oil shear rate is U/h and the force required to push the slider at a constant speed is

$$F = \mu(U/h)A$$

where A is the area of the slider in contact with the oil film.

When relative motion exists, the combination of a friction force and the motion results in the dissipation of energy so that work is done on the sliding bodies or on the fluid.

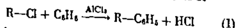
Friction is desirable in a number of cases, notably [illegible]

CHINE. [R.M.PH.]

Friedel-Crafts reaction

A substitution reaction, catalyzed by aluminum chloride, in which an alkyl (R—) group or an acyl (RCO—) group replaces a hydrogen atom of an aromatic nucleus. This general reaction is the most important member of a larger group of aromatic substitution reactions known to be catalyzed by conventional or Lewis acids.

Alkylation. In the classical alkylation reaction an alkyl halide (RX) serves as the alkylating agent. In the example in Eq. (1), an excess of



$$R{-}Cl + C_6H_6 \xrightarrow{AlCl_3} R{-}C_6H_5 + HCl \qquad (1)$$

the hydrocarbon, benzene (C_6H_6), may serve as the solvent. In case the aromatic hydrocarbon is not so readily available, a smaller quantity may be used in an inert solvent such as carbon disulfide or petroleum ether.

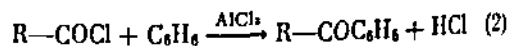
Alkylations usually require less than one mole of aluminum chloride per mole of alkyl halide, and a molar ratio of 1:10 is common. Tertiary halides (R_3CX) are superior to secondary (R_2CHX) or primary (RCH_2X) halides for the alkylation reaction. Rearrangement of the product may occur if the alkyl group is larger than ethyl (C_2H_5). For example, with *n*-propyl chloride ($CH_3CH_2CH_2Cl$), a large part of the product would be isopropylbenzene, $C_6H_5CH(CH_3)_2$, rather than the expected *n*-propylbenzene



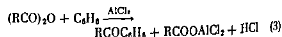
$$C_6H_5CH_2CH_2CH_3$$

Since alkyl groups attached to an aromatic nucleus facilitate further substitution, it is difficult to prevent di- or polysubstitution during the reaction. The structure of the dialkylation and trialkylation products cannot be predicted simply from a knowledge of the rules of substitution, since aluminum chloride is a catalyst for the intermolecular and intramolecular migration of alkyl groups. Alkenes (C_nH_{2n}) may be substituted for alkyl halides in the Friedel-Crafts reaction, and the industrial synthesis of ethylbenzene and isopropylbenzene (cumene) has been carried out in this way.

Acylation. For acylation of aromatic hydrocarbons, acyl halides, Eq. (2), have proved most valuable although acid anhydrides, Eq. (3), have also been used. For each mole of acid chloride present,



$$R{-}COCl + C_6H_6 \xrightarrow{AlCl_3} R{-}COC_6H_5 + HCl \qquad (2)$$



$$(RCO)_2O + C_6H_6 \xrightarrow{AlCl_3} RCOC_6H_5 + RCOOAlCl_2 + HCl \qquad (3)$$

slightly more than one mole of aluminum chloride is commonly required, whereas with acid anhydrides slightly more than two moles of aluminum chloride are necessary. The acyl group deactivates the benzene ring toward further substitution. Therefore, it is usually easy to obtain good yields of monosubstitution products. In contrast to the alkyl group, the substituted acyl group does not undergo rearrangement, making acylation a more reliable synthetic tool than the alkylation reaction.

Modifications of the Friedel-Crafts reaction are of value in the acylation of certain heterocyclic compounds such as thiophene and furan. Modifications have also been used in bringing about cyclization reactions through intramolecular acylation. *See* AROMATIC HYDROCARBON; BENZENE; SUBSTITUTION REACTION. [C.K.B.]

Bibliography: C. A. Thomas, *Anhydrous Aluminum Chloride in Organic Chemistry*, 1941.

Frigate bird

Any of five species of the family Fregatidae, sometimes called the man-of-war bird and found throughout the warmer seas of the world. Although it does not nest in Florida, one species, *Fregata magnificens*, the magnificent frigate bird, is frequently seen there. This graceful bird is 40 in. long and has a wing spread of 90 in. It is readily identified by its long, narrow wings and deeply forked tail.

Frigate birds are gregarious, feeding in groups and nesting in colonies. Although they can soar as



The frigate bird, *Fregata magnificens*; length to 41 in. (From P. Martin Duncan, ed., Cassell's Natural History, Cassell & Company, Ltd., London)

well as any birds known, they are not pelagic and usually stay near the shore. See PELECANIFORMES. [J.D.B.]

Fringe (optics)

A light or dark band that is produced by interference or diffraction of light. Fringes are usually very small, because of the short wavelength of light, and almost always require a magnifier for observation. They are clearer and more numerous when produced with light of a single color.

Diffraction fringes are formed when light from a point source, or from a narrow slit, passes by an opaque object of any shape. The Fraunhofer diffraction fringes produced when the light approaches and leaves the obstacle in essentially plane waves are especially important in the theory of optical instruments. See DIFFRACTION; RESOLVING POWER (OPTICS).

Interference fringes are obtained by bringing together two or more beams of light that have originated from a common source. This is usually accomplished by means of an apparatus especially designed for the purpose called an interferometer, although interference fringes may also be seen in nature. Examples are the colors in a soap film and in an oil film on water. When the fringes are controllable, for example, by changing the paths traversed by the two beams in an interferometer, they are valuable for the accurate measurement of small distances and of slight differences in refractive index. See INTERFERENCE OF WAVES; INTERFEROMETRY. [F.A.J.]

Fringing reef

A coral reef or other organic reef that fringes the edge of the land. A fringing reef is ordinarily divided into a steeply descending seaward front and a flat, broad, or narrow pavementlike surface that is awash at low tide. Although surfaces and fronts of such reefs may show vigorous growth of algae, corals, or other lime-secreting organisms, such growth is commonly only a veneer over an erosional sea-level bench. Charles Darwin's idealized sequence of reef development begins with the fringing reef, but studies of fossil and recent reef development show that it need not precede or be followed by more complex reef types. See CORAL REEF; ORGANIC REEF. [P.E.C.]

Frit

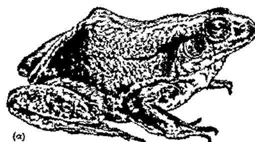
A ground glass used in making glazes and enamels. It is formed by melting the ingredients and rapidly cooling (quenching) them, for example, by pouring into water or onto water-cooled steel rolls.

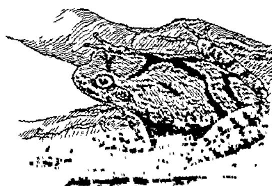
The main reason for fritting is to convert water-soluble raw materials into an insoluble glass, making it easier to keep them uniformly distributed in the glaze or enamel during processing. Other advantages of fritting are the conversion of poisonous materials (for example, lead compounds) into less dangerous forms, the more even distribution of color, and the reduction in loss of volatile materials.

The common practice is for the ceramic manufacturer to purchase frit from a large producer, thus achieving the additional advantages of a uniform raw material (and the resulting uniformity of processing), and a smaller inventory. See GLAZING; METAL COATINGS [M.C.M.]

Frog

Any of a large number of amphibians of the order Salientia (also called Anura), world-wide in distribution, and not sharply distinguished from the toads





Co... (a) The bullfrog, *Rana catesbeiana*; length of male to 7½ in. (b) The peeper, *Hyla crucifer*; length to 1¾ in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

—some animals commonly called toads and some called frogs are classified in the same suborder.

By far the great majority of the North American frogs belong to two families, the true frogs, or Ranidae, and the tree frogs, family Hylidae. The Leptodactylidae is represented by only a few species.

Typical frogs are smooth-skinned, relatively long-legged animals which stay in or near water or a damp habitat. Their tadpoles are slow in developing, usually waiting 1–2 years before transforming. In the colder limits of its range, the bullfrog may require 3 years to complete larval development.

Importance. Many frogs are used as food, and in the United States the rear legs from the larger species of *Rana* are considered a delicacy. Thousands of frogs of various species are used annually in high school and college biology classes, and many others are used in research in the medical and biological fields. One of the most common tests for pregnancy utilizes frogs, a use which requires many thousand frogs each year.

Kinds. In the United States the family Ranidae is represented by 18 species, all in the genus *Rana*. Best known and most abundant is the leopard frog, *Rana pipiens*, found in all regions except the Pacific Coast. This is a medium-sized frog with a body length of about 4 in. It is highly variable in color and several subspecies are recognized, but it is usually brown or green, marked with dark spots. This frog is widely used in research and medical laboratories.

The largest of the United States frogs is *Rana catesbeiana*, with a body length of about 7 in. This frog is usually dark green, but sometimes it is brown or black. It is found throughout the eastern United States and has been widely introduced in the West. It is the favorite food frog and is commonly used for dissection in zoology classes.

The tree frogs of the family Hylidae are found on all the continents. In the United States there are three widely distributed genera, and another, the least tree frog, is limited to the southeastern area. Best known is the common tree frog, *Hyla versicolor*, famous for its ability to change color and for its powers of "ventriloquism," its call seeming to come from a source other than the frog. The *Hyla* frogs have well-developed disks on their toes which help them to cling to objects. The spring peeper, *Acris gryllus*, and the chorus frogs of the genus *Pseudacris* are the animals heard peeping at the first sign of spring. *See* AMPHIBIA. [J.D.B.]

Front

A sloping surface of discontinuity in the troposphere, separating air masses of different density or temperature. Fronts are displaced by the wind, and the passage of a front at a fixed location is marked not only by sudden changes in temperature and wind but also by rapid variations in other weather elements, such as moisture, visibility, and sky condition. Because of the abrupt weather changes associated with them, fronts are major components in the analysis of surface weather maps and indispensable aids to accurate weather prediction. *See* AIR MASS.

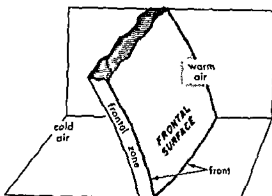


Fig. 1. Schematic diagram of the frontal zone, angle with earth's surface much exaggerated.

Although the front is ideally regarded as a discontinuity in temperature, in practice the temperature change from warm to cold air masses occurs over a zone of finite width, called a transition or frontal zone. The three-dimensional structure of the frontal zone is illustrated in Fig. 1. In typical cases the zone is about 3000 ft (1 km) in depth and 100 miles (100–200 km) in width, with a slope of approximately 1/100. The cold air lies beneath the warm in the form of a shallow wedge. The zone may extend from the earth's surface to the tropopause (35,000 ft), though generally the structure becomes diffuse above 20,000 ft. *See* ATMOSPHERE.

The surface separating the frontal zone from the adjacent warm air mass is referred to as the frontal surface, and it is the line of intersection of this surface with a second surface, usually horizontal or vertical, that strictly speaking constitutes the front. According to this more precise definition, the front represents a discontinuity in temperature gradient rather than in temperature itself. The boundary on the cold air side is often ill-defined, especially near the earth's surface, and for this reason is not represented in routine practice. In typical cases about one-third of the temperature difference between the Equator and the pole is contained within the narrow frontal zone, the remainder being distributed within the warm and cold air masses on either side.

The wind gradient, or shear, like the temperature gradient, is large within the frontal zone and discontinuous at the boundaries. An upper-level jet stream normally is situated above the zone, the strong winds of the jet inclining downward along or near the warm boundary. *See* JET STREAM.

Frontal waves. Many extratropical cyclones initiate as wavelike perturbations of a preexisting frontal surface. Such cyclones are referred to as wave cyclones.

The life cycle of the wave cyclone is illustrated in Fig. 2. In stage I, prior to the development, the front is gently curved and more or less stationary.

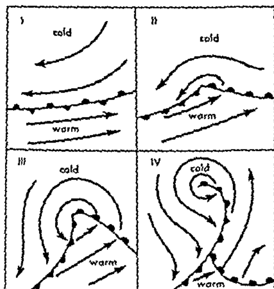


Fig. 2. The life cycle of the wave cyclone, surface projection. Arrows denote air flow.

In stage II the front undergoes a wavelike deformation, the cold air advancing to the left of the wave crest and the warm air to the right. Simultaneously a center of low pressure and of counterclockwise wind circulation appears at the crest. The portion of the front which marks the leading edge of the cold air is called the cold front. The term warm front is applied to the forward boundary of the warm air. During stage III the wave grows in amplitude, and the warm sector narrows. In the final stage the cold front overtakes and merges with the warm front, forming an occluded front. The center of low pressure and of cyclonic rotation is found at the tip of the occluded front, well removed from the warm air source. At this stage the cyclone begins to fill and weaken. See STORM.

A front moves with approximately the speed of the wind component normal to it. The strength of this component varies with season, place, and individual situation but generally lies in the range of 0–50 mph; 25 mph is a typical frontal speed.

Cloud and precipitation types and patterns bear characteristic relationships to fronts, as depicted in Fig. 3. These relationships are determined mainly by the vertical air motions in the vicinity of the frontal surfaces. Since the motions are not unique but vary somewhat from case to case and, in a given case, with the stage of development, the features of the diagram are subject to considerable variation. In general, though, the motions consist of an upgliding of the warm air above the warm frontal surface, a more restricted and pronounced upthrusting of the warm air by the cold front, and an extensive subsidence of the cold air to the rear of the cold front. See CLOUD; CLOUD PHYSICS.

The cloud and precipitation systems are narrower for fast-moving cold fronts than for slow, the main weather activity often running ahead of the cold front in the form of prefrontal squalls or squall lines. An occluded front contains elements of the cloud systems of both warm and cold fronts. See SQUALL.

Polar front. A front separating air of tropical origin from air of more northerly or polar origin is referred to as a polar front. Frequently only a fraction of the temperature contrast between tropical and polar regions is concentrated within the polar frontal zone, and a second or secondary front appears at higher latitudes. In certain locations such a front is termed an Arctic front.

In winter the major or polar frontal zones extend from the northern Philippines across the Pacific Ocean to the coast of Washington, from the southeastern United States across the Atlantic Ocean to southern England, and from the northern Mediterranean eastward into Asia. An Arctic frontal zone is located along the mountain barriers of western Canada and Alaska. In summer the average positions of the polar frontal zones are farther north, the Pacific zone extending from Japan to Washington and the Atlantic zone from New Jersey to the British Isles. In addition to a northward-displaced polar front over Asia, an Arctic front lies along the northern shore and continues eastward into Alaska.

Frontogenesis. The formation of fronts, or frontogenesis, has been studied mainly by considering the kinematic processes which give rise to increased temperature gradients. Confluent air currents which bring air from widely different source regions into juxtaposition are especially effective in this respect. The dissolution of fronts, or frontolysis, may be brought about by diffluent air motions and also by radiative, mixing, and other nonconservative processes.

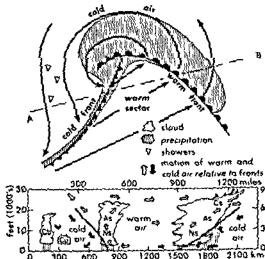


Fig. 3. Top: surface weather map. Bottom: vertical cross section (along A-B in top diagram). Cloud types: Cs, cirrostratus; As, altostratus; Ns, nimbostratus, Cu, cumulus; Cb, cumulonimbus.

Although frontogenesis can be described in terms of kinematic processes, there is as yet no basic explanation for the existence of fronts which has met with general acceptance. Laboratory experiments with heated, rotating fluids display frontlike phenomena under proper conditions of heating and rotation, so that it would appear that the front is a property of fluids generally and not of the atmosphere alone. [R.J.RE.]

Bibliography: C. L. Godske, et al, *Dynamic Meteorology and Weather Forecasting*, Carnegie Inst. Wash. Publ. 605, 1957.

Frost

A covering of ice in one of several forms produced by the freezing of supercooled water droplets on objects colder than 32°F The partial or complete killing of vegetation, by freezing or by temperatures somewhat above freezing for certain sensitive plants, also is called frost. Air temperatures below 32°F sometimes are reported as "degrees of frost"; thus 10°F is 22 degrees of frost.

Frost forms in exactly the same manner as dew except that the individual droplets that condense in the air a fraction of an inch from a subfreezing object are themselves supercooled, that is, colder than 32°F. When the droplets touch the cold object they freeze immediately into individual crystals. When additional droplets freeze as soon as the previous ones are frozen, and hence are still close to the melting point because all the heat of fusion has not been dissipated, amorphous frost or rime results.

At more rapid rates of condensation, the drops form a film of supercooled water before freezing, and *glaze* or *glazed* frost ("window ice" on house windows, "clear ice" on aircraft) generally follows. Glaze formation on plants, buildings, and especially wires sometimes is called an ice storm, or a silver frost storm, or thaw.

At slower deposition rates, such that each crystal cools well below the melting point before the next joins it, true crystalline or hoar frosts form. These include fernlike assemblages on snow surfaces, called *surface hoar*; similar feathery plumes in cold buildings, caves, and crevasses, called *depth hoar*; and the common window frost or ice flowers on house windows.

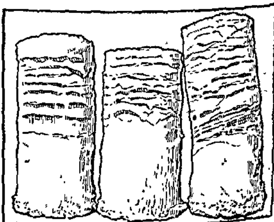
So-called killing frosts occur on clear autumn nights, when radiative cooling of ground, air, and vegetation causes plant fluids to freeze. At such times, the air temperature measured in a shelter 5–7 ft above ground usually is at least 5F° below freezing, but such standard level temperatures are poor indicators of frost severity. Air temperature varies greatly with height in the first few feet above the ground, and also with topography and vegetation around the shelter.

When wind is absent, the air layer immediately above the ground, rather than the ground itself, loses the most heat by radiation. Then the lowest temperature is 2–6 in. above, and about 1F° colder than, a bare ground surface. Above this near-ground minimum, an inversion of temperature develops for several to many feet thick. Plants radiate their heat faster than the air or ground, and may be colder than this air minimum temperature.

Frost damage can be prevented or reduced by heating the lowest air layers, or by mixing the cold surface air with the warmer air in the inversion above the tops of plants or trees. Cooled air becomes more dense, and tends to flow downhill. Thus valley bottoms are much colder on clear nights than slopes, and have frost when the slopes are frost-free. Frost pockets or hollows have been found in various parts of the world in places where the air temperature on some nights is many degrees colder than that at nearby weather stations situated on higher ground. *See* AIR TEMPERATURE; CLOUD PHYSICS; DEW; DEW POINT; PRECIPITATION (METEOROLOGY); TEMPERATURE INVERSION; WEATHER MODIFICATION. [A.C.]

Frost action

A term loosely used to indicate any action in which frost is involved. It has been defined by the American Geological Institute (1957) as the weathering process caused by repeated cycles of freezing and thawing. Engineers and botanists in particular include any effects of freezing and thawing, whether direct or indirect, within the scope of the term. Because of the multiplicity of meanings of the term, many geologists avoid its usage except informally and have introduced more specific terms. Some common self-evident subdivisions are frost-splitting, -wedging, -heaving, -thrusting, and -stirring. Frost action has been recognized for many decades as being exceedingly important in weathering and erosion, in disturbing vegetation, and in engineering problems. Frost action is intensified in the zone of seasonal near-surface freezing and thawing above perennially frozen ground (permafrost) and has been more widespread in the geologic past than it is today. *See* PERMAFROST; TUNDRA.



Ice lenses in frozen soil cylinders (*After S. Taber in D. P. Krynine and W. R. Judd, Principles of Engineering Geology and Geotechnics, McGraw-Hill, 1957*)

Frost action is initiated during growth of ice in the ground from which may result either hydrostatic pressure or directed pressure by individual ice crystals, usually growing in the direction of heat loss. Nine per cent volumetric expansion accompanies the transition of water to ice. In fine-grained soils, frost heaves are much greater and result from growth of ice lenses. The ice grows slowly as water migrates molecule by molecule through "tension" to the freezing zone The process is complicated and not entirely understood. Many factors are involved, such as the physicochemical nature of the material and its permeability, porosity, and rate of freezing. Plant roots are broken, soils are stirred, and many small related features are produced. In fine-grained unconsolidated materials, ice much in excess of the volume of pores is commonplace. Under these conditions thawing yields suspensions of soil or high pore-water pressure that promotes slump, flow, and loss of strength. See GEOLOGY; MASS WASTING; SOLIFLUCTION; WEATHERING PROCESSES. [R.F.B.]

Bibliography: R. F. Black, Permafrost, a review, *Bull. Geol. Soc. Am.*, 65:839-856, 1954; K. Bryan, Cryopedology, *Am. J. Sci.*, 224:622-642, 1946; A. W. Johnson, *Frost Action in Roads and Airfields*, Highway Research Board Spec. Rept. 1, 1952; K. B Woods et al., *Frost Action in Soils*, Highway Research Board Spec. Rept. 2, 1952.

Froude number

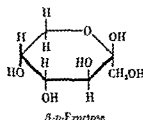
In fluid mechanics, the ratio v^2/gd of the inertia force v^2 to the gravitational force gd where v is velocity, g is gravitational acceleration, and d is a characteristic length. Froude number is significant in the design of a model of any system in which the effect of gravity is important in controlling the velocities or the flow pattern. For example, it is used in the evaluation of the drag or the slip-stream velocity of a ship producing surface waves. If the model is operated in the same gravitational field as the prototype, the similarity requirement based on equal Froude numbers in model and prototype leads to

$$v_m^2/d_m = v^2/d \quad \text{or} \quad v_m = v/\sqrt{n}$$

in which $n(d/d_m)$ is the length scale between model and prototype. See DYNAMIC SIMILARITY; MODEL THEORY. [C.M.]

Fructose

A sugar that is the commonest of ketoses and the sweetest of the sugars It is also known as D-fructose, D-fructopyranose, and levulose fruit sugar. It is found in free state, usually accompanied by D-glucose and sucrose in fruit juices, honey, and nectar of plant glands. D-Fructose is the principal sugar in seminal fluid. It can be isolated as crystalline β-D-fructopyranose, melting point (mp) 102–104°C It is strongly levorotatory in solution, having a specific rotation $[\alpha]_D = -132.2°$, which changes by mutarotation to $-92.2°$ (in water). Its



β-D-Fructose

course of mutarotation indicates the presence of an appreciable amount of the furanose form in solution. See CARBOHYDRATE; OPTICAL ACTIVITY.

The hydrolysis of cane sugar, or sucrose leads to the production of equimolar amounts of D-fructose and the D-glucose. On the other hand, inulin, a polysaccharide occurring in roots of the dahlia, chicory, and many Compositae, yields D-fructose alone. From the latter sources, it is prepared industrially.

As represented in the above formula, fructose normally occurs as fructopyranose, with a 1,5 oxidic ring. In combined form, however, as in sucrose or inulin, the fructose residue contains the 1,4 ring, and thus is of the fructofuranose type.

D-Fructose is more soluble in water than D-glucose and has a much sweeter taste. It is readily fermented by bakers' yeast, resulting in the same products, ethyl alcohol and carbon dioxide, as glucose.

D-Fructose is readily utilized by diabetic animals. In persons with diabetes mellitus or parenchymal hepatic disease, the impairment of fructose tolerance is relatively small and not at all comparable to the diminution in their tolerance to glucose. See METABOLIC DISORDERS; MONOSACCHARIDE. [W.Z.H.]

Fruit (botany)

A fruit is a fully matured ovary with or without other floral or shoot parts united with it at maturity. The ovary develops into a fruit after the egg is fertilized and usually contains one or more seeds. See SEED (BOTANY). The fruit is significant to the plant with regard to seed dissemination. The type of fruit is determined by the structure of the ovary and its relation to the other parts of the flower. The wall of the fruit, developed from the wall of the ovary, is called the pericarp. The pericarp may be dry or fleshy, and the latter type may range from soft and juicy to hard and tough.

Kinds of fruits. Dry fruits may be dehiscent, opening for release of seeds, or indehiscent, remaining closed and usually containing but one seed. The lines of opening of dehiscent fruits are called sutures and the separating segments are valves (Fig. 1). Dehiscent fruits, often uncritically called pods, are represented by the following types: follicle, composed of one carpel (the basic unit of an ovary) opening along one suture, as in milkweed; legume, with one carpel but opening along two sutures, as in common bean; loment, a legume, constricted between seeds and breaking crosswise into one-seeded segments, as in sticktights; silique,

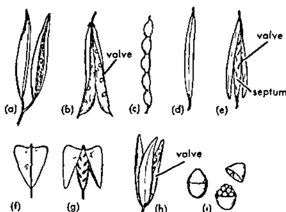


Fig. 1. Dehiscent dry fruits. (*a*) Follicles of milkweed. (*b*) Legume (bean). (*c*) Loment (sticktights). (*d*) Silique (mustard) closed (*e*) Silique (mustard) open. (*f*) Silicle of shepherd's purse, closed. (*g*) Silicle of shepherd's purse, open (*h*) Capsule (lily). (*i*) Pyxis (purslane).

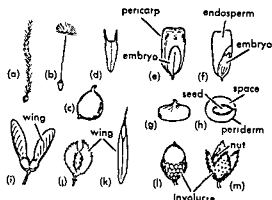


Fig. 2. Indehiscent *dry fruits.* (*a*) Achenes of clematis. (*b*) Achenes of dandelion. (*c*) Achenes of buttercup. (*d*) Achenes of beggarticks. (*e*) Caryopsis of corn. (*f*) Vertical section of e. (*g*) Utricle of goosefoot. (*h*) Section of g to show space surrounding seed. (*i*) Samara of maple. (*j*) Samara of elm. (*k*) Samara of ash. (*l*) Nut (acorn) of oak. (*m*) Nuts of beech.

composed of two *elongated* carpels with two sutures, opening from below upward exposing the seed-bearing partition (septum) still attached, as in mustard; silicle [illegible]

[illegible] as in lily. There is much variation, however, in the manner of dehiscence of a capsule. For example, in a type called pyxis the dehiscence is circumferential, the top coming off like a lid, as in purslane.

Ind[illegible]

ca[illegible]

ca[illegible]

co[illegible] utricle, like an achene but with pericarp thin and bladdery, as in goosefoot; samara, like an achene but provided with a wing for wind dispersal, as in maple; nut, larger than achene with pericarp hard throughout, usually enclosed or partly so in an involucre, as in acorn and hazelnut (Fig. 2).

Among fleshy fruits, three general groups prevail: simple, developed from a single ovary, as in peach; aggregate, derived from numerous ovaries of a single flower, as in blackberry; multiple, including ovaries and associated structures of several flowers consolidated into one mass, as in mulberry and pineapple (Fig. 3).

Of simple fleshy fruits, the chief types are pome, having a cartilaginous endocarp (core) with the outer pericarp and thickened floral tube forming the fleshy part of fruit, as in apple; drupe (stone fruit), with the pericarp differentiated into three distinct layers, exocarp (skin), mesocarp (fleshy portion), and endocarp (inner stony layer), as in cherry; berry, in which the entire ovary becomes pulpy, as in tomato. A modified berry, such as orange, with a thick leathery separable rind containing numerous oil glands is called a hesperidium, whereas one with a hard or tough inseparable rind is a pepo, as in cucumber. [N.A.]

Fruit anatomy. This area of morphology is concerned with the internal structure of fruits. A fruit develops from the basic unit of the gynoecium (female part of the flower), the carpel, or from several more or less closely united carpels. Usually only part of the gynoecium, the ovary, develops into a fruit; the style and stigma wither. Accessory (extracarpellary or noncarpellary) structures may be closely associated with the carpel (or carpels), displaying various degrees of fusion to them (adnation), and thus becoming part of the fruit. Such noncarpellary parts are, for example, sepals (mulberry), bases of sepals, petals, and stamen united into a floral tube (banana, apple), the receptacle (strawberry), the peduncle (fleshy part of fig), the pedicel (cashew), the involucre comprised of bracts and bracteoles (walnut, pineapple), and inflorescence axis (pineapple). *See* FLOWER (BOTANY).

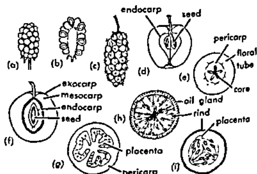


Fig. 3. Fleshy fruits. (*a*) Aggregate (blackberry). (*b*) Vertical section of *a*. (*c*) Multiple (mulberry). (*d*) Pome (apple), vertical section. (*e*) Pome (apple), cross section. (*f*) Drupe (cherry), vertical section. (*g*) Berry (tomato), cross section. (*h*) Hesperidium (orange), cross section. (*i*) Pepo (cucumber), cross section.

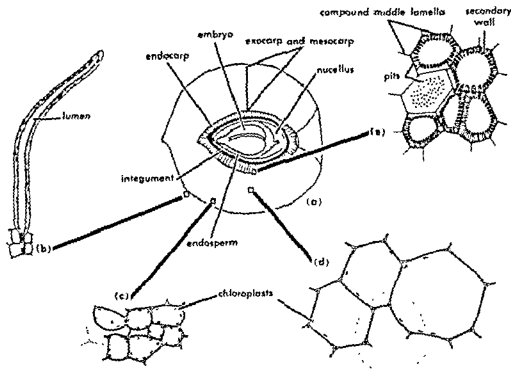


Fig. 4. Tissue elements of the fruit of Prunus (peach). (a) Diagrammatic drawing of a longitudinal section of peach fruit. The rectangles accompanied by small letters indicate the positions of the sections shown in b, c, d, and e. (b) Epidermal hair. (c, d) Parenchyma of the mesocarp. (e) A group of sclereids from the endocarp. (From K. Esau, Plant Anatomy, Wiley, 1953)

A fruit derived from carpellary structures alone (Fig. 4) is called a true fruit (rose, date, tomato, peach, corn), or because it arises from a superior ovary (inserted above the other floral parts), a superior fruit. Fruits in which accessory tissues are included are called accessory, false, spurious, or because of derivation from inferior ovaries (inserted below the other floral parts), inferior fruits (apple, walnut, squash, coconut). True fruits develop from hypogynous or perigynous flowers (superior ovary), false fruits usually from epigynous or semiepigynous ones (inferior ovary). Exceptions occur as in strawberry which is an accessory fruit because of the large receptacle bearing the small achenes (true fruitlets), yet the flower from which it develops is hypogynous. True and accessory fruits may be distinguishable from each other on the basis of developmental anatomy.

Fruit wall. In true fruits, the fruit wall is synonymous with pericarp. The fruit wall in accessory fruits (Fig. 5) includes the pericarp plus one or more accessory tissues of various derivations. At maturity, fruit walls may be fleshy, composed primarily of parenchyma, the cells of which may be closely appressed or loosely arranged with intercellular spaces, or dry, often including sclerenchyma (fibers and stone cells), or a combination of both types forming distinct layers, as in the peach, or with the cell types intermingled, as in the pear (see PARENCHYMA; SCLERENCHYMA). In epigynous fruits, the anatomical distinction between pericarp and floral tube is not always clear (lettuce). However, in certain apple cultivars, for example Pink Pearl, the pericarp and accessory tissues of the fruit wall are distinguished by color, the pericarp white and the accessory tissues pink. Tannins, starch grains, latex, and plastids are regular inclusions in the cells of many fruits. Anatomically some fruit walls, such as in sunflower (achene) and grains (caryopsis), resemble seed coats. The identity of cell layers present at fertilization or during early fruit development may become obscured later. All cells of the mature fruit wall may be present at fertilization, with fruit growth resulting only from cell enlargement; or the fruit may develop by cell division, which may continue briefly (peach) or until maturity (avocado, apple), accompanied by cell enlargement. Russeting occurs in apples, pears, and sapote when cork develops in some of the outer cell layers

Pericarp. The wall of the carpel (or carpels) develops into the pericarp, an essential part of any fruit. In false fruits, there is always a pericarp present as well as other structures (all termed fruit wall). The pericarp is usually divisible anatomically into three separate parts: exocarp, mesocarp, and endocarp. In fleshy fruits, cells are soft and juicy (tomato) or parts of the fruit are soft and others dry (peach). Cells of soft types may separate at maturity (jujube, watermelon) or may not sepa-

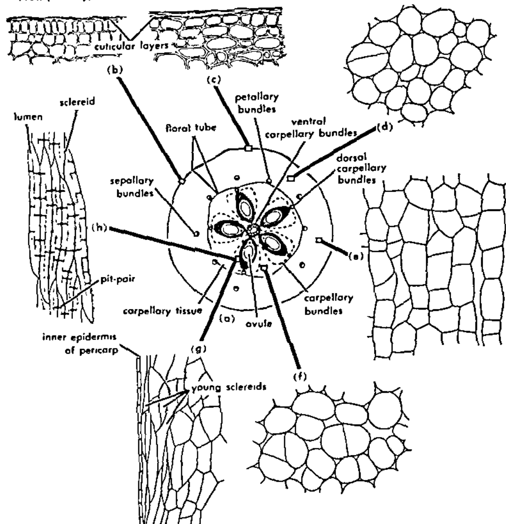


Fig. 5. Tissue elements of the fruit of *Malus* (apple). (*a*) Diagrammatic drawing of a transection of apple fruit. The rectangles accompanied by small letters indicate the positions of sections shown in *b*, *c*, *d*, *e*, *f*, *g*, and *h*. (*b*, *c*) Epidermis and subjacent collenchymatous tissue from young and mature fruits. (*d*, *e*) Parenchyma of the floral-tube part of the flesh. (*f*) Parenchyma from the exocarp. (*g*) Endocarp from young fruit. (*h*) Endocarp from mature fruit. (*From K. Esau, Plant Anatomy, Wiley, 1953*)

rate (persimmon, grape, sour cherry). Dry fruits are usually more complex than fleshy ones, with various intermixing of parenchyma and sclerenchyma, perhaps the latter with two or more layers oriented in different directions. Variable drying rates and differential anatomical weaknesses aid in causing dehiscence of some dry fruits (fleshy fruits rarely dehisce). Structures other than the pericarp may compose a principal portion of a fruit, for example, the expanded placenta in tomato, and the mucilaginous seed coats of pomegranate.

Exocarp. The exocarp, sometimes called epicarp, may merge imperceptibly with the outer fruit wall of accessory fruits as in apple. In true fruits, the exocarp may be very thick and distinct (skin of plum) or it may consist only of epidermis (pea). It may be smooth as in the avocado, or it may have numerous unicellular hairs as in the apricot (*see* EPIDERMIS, PLANT). Stomata may be present (citrus) or lacking (grape). There is usually a hypodermal region of isodiametric cells which may extend from one to seven or more cell layers deep, and may contain anthocyanins, carotenoid pigments (*see* CAROTENOID), tannins, starch grains, collenchyma (*see* COLLENCHYMA), sclereids, or oil glands. In mature fruits, lenticels may replace stomata. Exocarps may be narrow or wide, composed of thin- or thick-walled cells, and be dry or fleshy. The external tissues of accessory fruits may have the same structure as exocarps of true fruits.

Mesocarp. Located between exocarp and endocarp, the ground tissue of the mesocarp may be a homogeneous parenchyma (peach flesh), or consist of parenchyma and sclerenchyma (sclereids, fibers) of varying complexity as in legumes, or parenchyma with islands of sclerenchyma (pear). As maturity approaches, some fleshy tissues become disorganized and show less wall thickening (peach).

The cells may be thin- or thick-walled, often with their long axes oriented at right angles to each other, so that a mechanically strong structure results. In some fruits there is no way to distinguish mesocarp from endocarp.

Endocarp. A single simple thin- or thick-walled epidermal layer next to the locule and seed may comprise the entire endocarp, but usually it is more complex. Inwardly, the endocarp may produce multicellular hairs which become juice vesicles (citrus). The endocarp may be distinct as in the highly sclerified stone of peach, or with layers of thick-walled cells oriented in two directions. Often the endocarp is parenchymatous and very similar anatomically to the mesocarp. Cuticle may be present on the inner epidermis.

Vascular system. The primary vascular bundles in fruits vary from one (strawberry achene) to five or more per carpel, with three the fundamental number—two adaxial carpellary bundles (ventral) which supply the seed (or seeds) and one abaxial (dorsal) which is found opposite the ventrals, across the locule. Cohesion of bundles is characteristic of many compound (syncarpous) fruits so that only one adaxial bundle is serving one carpel or two adaxials are serving two carpels. Vascular bundles anastomose (interjoin) outward to the inner part of the hypodermal tissue, thus forming a complete network in the entire fruit. Bundles may be the principal support of all other tissues in light-weight, feathery fruits (*see* PHLOEM; XYLEM).

Abscission. Flower parts, unless anatomically united with the carpel or carpels, usually fall during development of the fruit, with the fruit (ovary) being the only floral organ remaining at maturity. Fruits usually abscise (separate) from the plant when ripe (peach, coconut, *Platanus*). The fruit and its stalk may separate as a unit from the tree, as in apple, or the pedicel may remain on the branch when the fruit abscises, as in peach. An abscission zone may involve either the formation of corklike cells across the pedicel or at the base of the fruit, or only physiological changes in existing cells, especially in cell walls. Both steps lead to separation of fruit from the plant. Some cells show no changes but are easily broken mechanically at abscission. No abscission zone is observed in some species, yet the fruit will fall. [R.M.BR.]

Bibliography: See PLANT ANATOMY.

Fruit (tree)

Tree fruits include temperate-zone species, citrus, subtropical, and tropical fruits. The approximate annual values of the first three classes in the United States are shown in the table.

Temperate-zone fruits. The principal fruit trees in the temperate zone are deciduous, that is, they lose their leaves in the fall of the year (*see* DECIDUOUS PLANTS). They are cultivated principally in regions protected from prolonged summer heat and severe winter cold (above −10 to −15°F).

Approximate annual value of United States fruit crop

Temperate-zone fruits	
Apple	$222,000,000
Peach	133,000,000
Pear	61,330,000
Plum	60,700,000
Cherry, sweet	26,970,000
Cherry, sour	21,400,000
Apricot	21,000,000
Nectarine	5,300,000
Citrus fruits	
Orange	223,000,000
Grapefruit	43,469,000
Lemon	40,811,000
Tangerine	8,500,000
Lime	1,135,000
Subtropical fruits	
Grape (all kinds)*	130,000,000
Olive	9,000,000
Avocado	8,500,000
Fig	8,500,000
Date	2,000,000

* Approximately 80% produced in California

The apple, *Malus domestica*, is the most cosmopolitan of the deciduous tree fruits, leading in production both in the United States and in the rest of the world. The United States production (short tons) in 1957 of the major deciduous tree fruits was apple, 2,386,500 (world 11,825,000); peach, *Prunus persica*, 1,513,000; pear, *Pyrus communis*, 780,000; plum, *Prunus domestica* and *P. salcina*, 581,000; apricot, *Prunus armeniaca*, 208,000; sour cherry, *Prunus cerasus*, 149,500; sweet cherry, *Prunus avium*, 89,000; and nectarine, *Prunus persica*, 36,000. [H.B.T.]

Citrus fruits. Cultivated species of citrus fruits include a large number of varieties belonging to the genera *Citrus*, *Fortunella*, and *Poncirus*. All are closely related and belong to the subtribe Citrinae, tribe Citreae, subfamily Aurantioides, family Rutaceae, and order Geraniales (*see* GERANIALES). The commonly known fruits, including the citron, grapefruit, lemon, lime, orange, and tangerine, are in the genus *Citrus*, and are believed to be native to the subtropical and tropical regions of Asia and the Malay Archipelago.

The large number of varieties in the various species of citrus and the endless array of hybrids between species result in a group of plants quite diverse in vigor, appearance, and character of growth. However, all are evergreen trees or shrubs (*see* EVERGREEN PLANTS). The flowers are usually white and very fragrant and are borne singly or in short axillary racemes. *See* FLOWER (BOTANY). Botanically, the fruit is a berry. It has segments which, unlike any other fruits known to botanists, are filled with pulp vesicles (bladderlike structures) containing, in many varieties, a delicious juicy tissue. *See* FRUIT (BOTANY).

Practically all commercial citrus plantings are propagated asexually by budding the desired variety onto young seedlings of several different species which serve as rootstocks (*see* BUDDING; RE-

PRODUCTION, PLANT). Sour orange, sweet orange, rough lemon, Cleopatra mandarin, and *Poncirus trifoliata*, a near relative of true citrus, are most commonly used as rootstocks.

In the United States, which for a number of years has produced approximately 45% of the world supply of citrus fruits, commercial plantings are restricted almost entirely to Florida (527,600 acres), California (210,100), Texas (42,800), Arizona (13,800), and Louisiana (4,500). While these are approximate acreages as of 1958, the trend of citrus plantings is increasing in Florida, Texas, and Arizona and decreasing in California and Louisiana. The total value of the citrus crop delivered at the packing houses in the United States averaged $316,737,000 per year from 1948 to 1958.

Citrus, in the form of fresh fruit or juice, is the dietary fruit with the largest consumption in the United States and is the most important source of vitamin C (*see* VITAMIN). [F.E.C.]

Subtropical fruits. These are adapted to the climatic zones adjoining the tropics and separating the tropics from the temperate zones. In the United States this area extends from north central California through southern Arizona, the lower Rio Grande Valley of Texas, and across to and including Florida and south Georgia. Subtropical fruits are either evergreen, such as olive, citrus, avocado, and date, or deciduous such as fig and European grape. [C.A.S.]

Tropical fruits. The two main tropical fruits of commerce are the banana and date. The date is grown to a limited extent in hot desert areas of California and Arizona, but the banana has never been economically produced in the United States.

See separate articles for fruits listed under common names in this article. [J.F.F.]

Bibliography: *See* AGRICULTURAL SCIENCE (PLANT).

Fruit (tree) diseases

Tree fruits have diseases that may spoil the appearance of the fruits or render them unfit for human consumption; diseases may also lower yield or sh[illegible] tre[illegible] spe[illegible] dis[illegible] zor[illegible] called deciduous fruits) are due to fungi, bacteria, viruses, and adverse environmental conditions.

Fungus diseases. Fungi are the most frequent causes of temperate-zone tree-fruit diseases. For example, a perennial threat to most commercial and home apple orchards is the apple scab disease (Fig. 1). Caused by the fungus *Venturia inaequalis*, scab may cause premature defoliation, June drop of young fruits, and unsightly blemishes on ripe apples. A similar disease attacks most varieties of pears and is caused by *V. pyrina* (*see* FUNGI).

Stone fruits, such as peaches, plums, cherries, apricots, and nectarines, are frequently damaged

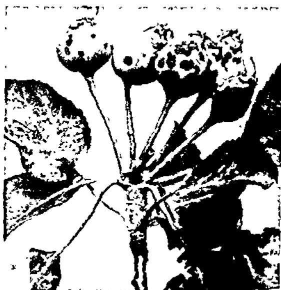


Fig. 1. Apple scab on young fruit.

extensively by brown rot caused by the fungus *Monilinia fructicola* (Fig. 2). This disease attacks the flowers (causing blossom blight), infects terminals and twigs with twig blight, and is widely recognized by its destruction of ripe fruit at harvest time.

Many diseases caused by fungi attack and destroy leaves of tree fruits: cherry leaf spot, *Coccomyces hiemalis*; peach leaf curl, *Taphrina deformans*; apple cedar rust, *Gymnosporangium juniperi-virginianae*, leaf blight of pears, *Mycosphaerella sentina*; and apple powdery mildew, *Podosphaera leucotricha* (Fig. 3).

Bacterial diseases. Bacteria also attack tree fruits, causing extensive damage in some regions. For example, fireblight of apples and pears, *Erwinia amylovora*, is a common disease of pomaceous fruits in most seasons (Fig. 4). Bacterial leaf spot of stone fruits, *Xanthomonas pruni*, is often the cause of serious damage to leaves, twigs, and fruits of peaches, apricots, almonds, cherries, plums, and nectarines. Crown gall, *Agrobacterium tumefaciens*, may infect tree fruits in the nursery row, causing the formation of large galls on roots and stems of young apple, peach, and pear trees. *See* BACTERIA.

Virus diseases. Diseases caused by viruses, such as yellows, rosette, and phony, are especially prevalent on peaches, plums, cherries, apricots, and almonds. *See* PLANT VIRUS.

Effect of environment. The growth of fruit trees and the production of fruit may also be affected by unfavorable environmental conditions such as infertile soil, low temperatures, drought, and poor drainage. [E.C.SH.]

Citrus diseases. Citrus is subject to a great number of diseases, most of which are distributed throughout the citrus-producing countries of the world. Infectious diseases of citrus are caused by bacteria, fungi, and viruses and may be grouped according to the parts of trees affected, that is, fruit, leaves and twigs, and trunks and roots.

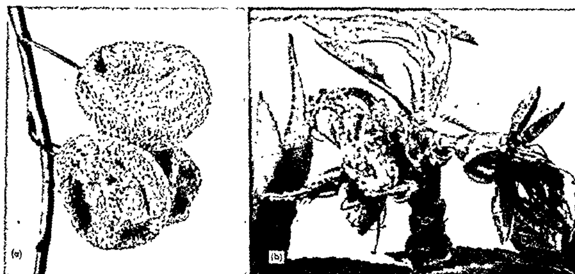


Fig. 2. (a) Plum brown rot. (b) Peach brown rot (blossom blight).

Fruit. Numerous rots, internal derangements, and external markings result from fungal and bacterial infections. Fruit decay develops chiefly during storage and transit, from infections established largely in surface wounds during growing, and from picking, packing, or marketing operations. Citrus fruit

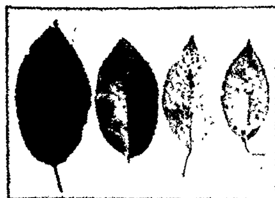


Fig. 3. Cherry leaf spot.

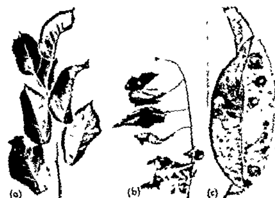


Fig. 4. (a) Healthy apple twig. (b) Fireblight-diseased apple twig. (c) Bacterial leaf spot of peach.



Fig. 5 Blue-green molds (*Penicillium italicum* and *P. digitatum*) take a heavy toll in storage, transit, and market. The illustration shows how these decay-causing molds spread by contact from an infected fruit. (Photograph by L. J Klotz)

requires careful handling to avoid economic loss from decay caused by such fungi as the common green mold, *Penicillium digitatum*, blue mold, *P. italicum* (Fig. 5); brown rot, *Phytophthora citrophthora*, and cottony-rot, *Sclerotinia sclerotiorum*. Some decays are described as contact rots because the causal fungi can spread through an entire box of fruit from one infected fruit.

Leaves and twigs Some of the fruit-rotting fungi and others cause leaf-blight and dieback of twigs. A bacterium, *Xanthomonas citri*, is the cause of citrus canker which attacks leaves, twigs, and fruits. Introduced on citrus plants sent from Japan to Florida in 1913, it threatened to destroy the citrus industry there. Its eradication, requiring destruction by burning of 4,000,000 orchard and nursery trees, is a most remarkable achievement in plant disease control.

Trunks and roots. Some of the fruit-rotting fungi, particularly species of *Phytophthora* and *Diplodia*, are responsible for root rots and gumming diseases of trees (Figs. 6, 7).

Virus diseases. Citrus is subject to several destructive virus diseases. One of these, psorosis, causes development of bark lesions and eventual death of trees. Since it was learned that psorosis is a bud-perpetuated disease that does not spread naturally, it has been possible to avoid it by propagating from psorosis-free sources. Another virus disease, tristeza, causes rapid decline or death of trees of sweet orange, grapefruit, and tangerine propagated on certain susceptible rootstock varieties. Tristeza virus is known to be spread by three species of aphids. It virtually destroyed the citrus industries of Argentina and Brazil in the 1930s. Damage from tristeza in the United States has been less because more of the citrus is on resistant rootstocks and the two species of the genus *Aphis* that spread it here are less efficient vectors than the one species common in South America. [J.M.WA.]



Fig. 6. Gummosis, a disease caused by the fungus *Phytophthora citrophthora*, damages or kills many citrus trees. Infection by this fungus, one of the water molds, is favored by excessive moisture. (*Photograph by L. J. Klotz*)

Bibliography: *See* PLANT DISEASE.

Fruit growing (small)

Cultivation of the small fruits which include those commonly known as berries, such as blackberries, dewberries, raspberries, cranberries, blueberries, strawberries, gooseberries, and currants. Grapes are occasionally included with the small fruits in horticultural publications, but more frequently are listed as vine fruits. There is no botanical category which includes all the small fruits because they represent diverse groups of at least three plant families and several genera. They are considered together because of their size, similarity of cultural requirements, and the fact that all, except the cranberry, are well adapted to culture in the home garden.

Cranberries are grown extensively in very limited localities in Massachusetts, Wisconsin, New Jersey, Washington, and Oregon, where marsh land and cool temperatures are available. Blueberries are confined primarily to limited areas in New Jersey, Michigan, North Carolina, and Washington. They grow best in well-drained, sandy, peat land, somewhat similar to cranberry land, but there is



Fig. 7. Original Washington Navel orange tree planted in Riverside, California, in 1873. This historical tree was threatened by an attack of *Phytophthora* gummosis in 1918, but was saved by inarching to seedling trees of gummosis-resistant varieties which now support it. The original root system was completely destroyed by the fungus infection. (*Photograph by L. J. Klotz*)

medium-sandy loams; blackberries, raspberries, and strawberries do well on sandy loam to clay loam; and currants and gooseberries on medium to clay loam. In a few cases growers raise several kinds of small fruit, but usually a grower will produce only one type of berry, frequently as a sideline in addition to other farm crops or tree fruits. One of the major problems associated with the production of all berries except the cranberry is that picking is by hand, which requires a rather large temporary labor force during the harvest season.

Aside from gooseberries, currants, and cranberries, the fruit of many berries is consumed fresh for dessert purposes, hence is sold on the fresh market to the extent that the available demand will absorb it; the remainder, which usually constitutes a major part of the crop, is processed for sale as frozen fruit, or is made into such products as jellies and preserves. The total small-fruits industry of the United States returns close to $100,-000,000 per year to the growers. Because of the different types of fruit and the scattered location of the plantings, there is no cohesive force to tie all

Some important small-fruit-producing states

Crop	Producing states
Strawberries	Oregon, California, Tennessee, Michigan, Louisiana, Washington, Arkansas, Kentucky, New York
Raspberries	Michigan, Oregon, New York, Washington, Ohio, Pennsylvania, Minnesota
Blackberries and dewberries	Oregon, Texas, California, Washington, Michigan, Arkansas, Oklahoma, Alabama, North Carolina
Currants and gooseberries	New York, Oregon, Washington, Michigan

the growers together into a nationwide industry, as there is in certain other branches of agriculture. See separate articles for important small fruits listed under their common names; *see also* FRUIT (TREE). [J.H.CE.]

Frustration

The experience of nonfulfillment of some wish or need. The concept of frustration is only beginning to play a systematic role in scientific behavior theory. In 1939 J. Dollard and his associates advanced the so-called frustration-aggression hypothesis, which held quite simply that when living organisms experience frustration they are instigated to aggression. However, it was quickly pointed out by S. S. Sargent among others, that individuals vary enormously in the way they handle frustration; or, to use S. Rosenzweig's phrase, they vary in their frustration tolerance. These and similar criticisms led N. E. Miller to reformulate the frustration-aggression hypothesis. Instead of holding that frustration is the necessary and sufficient condition for aggression, he changed the hypothesis to read, "frustration produces instigations to a number of different types of response, one of which is . . . aggression." *See* AGGRESSION.

Freudian concept. Interest in the problem of frustration along lines indicated above was largely inspired by the writings of Sigmund Freud. Freud postulated that the blocking, inhibition, or frustration of sexual and other instinctive impulses is the basic cause of the psychoneuroses; and since childhood discipline implies frustration, there was a widespread reaction, under the influence of Freudian thought, against such discipline. This reaction was, however, short-lived, for it was soon discovered that complete indulgence and unrestricted permissiveness in the child results in personality problems even more malignant than those which such practices were designed to correct. Today many psychiatrists, psychologists, and even psychoanalysts are no less strict with their children than the average parent tends to be. *See* PSYCHOSIS.

Frustration-extinction concept. Recently there has been an upsurge of interest in frustration theory along lines different from those thus far considered. Students of learning have long known that the withholding of accustomed reward extinguishes a habit, but there has been no generally accepted explanation of this phenomenon. In 1958 A. Amsel pointed out that nonappearance of an expected reward constitutes a frustration and that, under some conditions, frustration leads not to aggression but to extinction, that is, not to intensification but to a weakening of the frustrated response tendency. *See* CONDITIONED REFLEX; LEARNING THEORIES.

Intermittent reinforcement. The question of when frustration leads to anger and aggression and when to disappointment and extinction has not yet been well worked out, but one aspect of this new dimension of frustration theory appears to be of considerable importance. It has been found by L. G. Humphreys that a habit is more resistant to extinction if during acquisition, training reinforcement or reward is intermittent rather than continuous, that is, occurs only on occasional trials rather than on all of the acquisition trials. Many attempts have been made to explain this paradoxical effect, none of which has been entirely convincing. The difficulty seems to lie in the assumption that intermittent reinforcement produces greater habit strength than does continuous reinforcement and, hence, greater resistance to extinction. It now appears that intermittent reinforcement during acquisition accustoms the individual to frustration, in small doses, so that massive or continuous frustration, when encountered, in so-called extinction, is less effective. In other words, intermittent reinforcement seems to increase an organism's frustration tolerance, so that the experiences connected with the procedure of extinction or continuous nonrewarded trials are rendered less punishing than they are following acquisition which has involved continuous reward, that is, no prior experience with absence of reward and frustration.

Attempts to develop systematic theory concerning the frustration-aggression sequence have been relatively unsatisfactory. Much the same can be said for N. Maier's frustration-fixation hypothesis. Learning theorists seem to be on sounder ground in their analysis of the frustration-disappointment sequence and its implications for unlearning or extinction. *See* PSYCHOLOGY, PHYSIOLOGICAL AND EXPERIMENTAL. [O.H.M.]

Bibliography: A. Amsel, The role of frustrative nonreward in noncontinuous reward situations, *Psychol. Bull.*, 55:102–119, 1958; J. Dollard, et al., *Frustration and Aggression*, 1939; S. Freud, *A General Introduction to Psychoanalysis*, 1920; L. G Humphreys, Acquisition and extinction of verbal expectations in a situation analogous to conditioning, *J. Exptl. Psychol.*, 14:294–301, 1931; N. R. F. Maier, *Frustration: The Study of Behavior without a Goal*, 1949; N. E. Miller, The frustration-aggression hypothesis, *Psychol. Rev.*, 48:337–342, 1941; S. S. Sargent, Reaction to frustration: a critique and hypothesis, *Psychol. Rev.*, 55:108–114, 1948.

Fuel

A material whose combustion is used to supply heat for any purpose. Most fuels are carbon-containing materials that are burned with air or with

derived from the air. For a discussion of notable exceptions *see* CHEMICAL FUEL; METAL-BASE FUEL.

Classification of fuels. Fuels are characterized first by their physical form at normal temperatures, whether they are solids, liquids, or gases; second, by their heating value, that is, by the amount of heat given off when a unit weight or volume of the material is burned under standard conditions; and third, by their combustion characteristics. Combustion characteristics cover such points as ease of ignition, rate of combustion, flame temperature, and flame luminosity.

Fuels may be used to supply heat directly, as in a furnace. They may also be used to supply heat to perform some other function, such as to produce steam in a boiler, the steam then being used to produce power in a turbine or engine. The heat may be used to raise the temperature and pressure of the products of the combustion to provide a driving force, as in the internal combustion engine or gas turbine. *See* ENGINE; FURNACE (STEAM GENERATING); TURBINE.

Solid fuels today comprise primarily the various ranks of coal (anthracite, bituminous, subbituminous, and lignite) and the coke or char derived from them. Wood, charcoal, peat, and various other plant products, while important under special circumstances, represent only a small fraction of total fuel consumed. The important factors governing choice of coal for a given use are cost per unit of heating value, size range, moisture and ash content, and amount of smoke evolved. *See* COAL; COKE.

Liquid fuels are derived almost entirely from petroleum. Historically, coal oil (a kerosine derived from coal tar) was used extensively for heat and light, but it is now largely supplanted by petroleum products. In general, the liquid fuels suitable for internal combustion engines command a premium, and the fuels designed for external combustion are derived from those portions of the petroleum that have the least potential for gasoline and diesel-fuel manufacture. Factors in choice of liquid fuels are heating value, fluidity, boiling range, impurities such as sulfur, and freedom from water and sediment. For fuels for internal combustion engines, *see* DIESEL FUEL; FUEL OIL; GASOLINE; PETROLEUM PRODUCTS.

Of the gaseous fuels, the most important are those derived from natural gas or petroleum, chiefly methane. Availability of these has increased sharply since World War II and they have largely supplanted manufactured fuel gas for general distribution by public utilities. Manufactured gas was usually coke-oven gas or carbureted water gas made from coke, steam, and

Heating value of fuels. The heating value of gaseous fuels varies widely with composition. For this reason, it is generally important to have an accurate and continuous determination of the heating value, and much effort has been devoted to the development of an apparatus for this purpose. Carefully metered streams of gas and combustion air are burned under constant temperature conditions in a submerged chamber. The temperature rise of a metered stream of cooling air is used as a measure of the heating value and is usually recorded continuously on a chart. This value in the United States is usually expressed as British thermal units (Btu) per cubic foot of gas.

Approximate heating values of representative fuels

Fuel	Heating value
Producer gas	129 Btu/ft³
Butane	3,200 Btu/ft³

Heating values of solid and liquid fuels are determined by oxidation of a weighed sample in a system whose heat capacity is known and whose temperature rise can be measured. The oxidant may be gaseous oxygen or an oxidizing material such as sodium peroxide.

Heating values are reported as gross (or higher) heating value when the water formed in combustion is condensed, and as net (or lower) heating value when the water from combustion leaves the system as a vapor and the heat of condensation is not included. *See* CALORIMETRY; COMBUSTION; FURNACE CONSTRUCTION; SMOKE. [H.R.B.]

Bibliography: John Griswold, *Fuels, Combustion and Furnaces*, 1946.

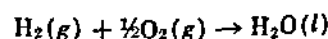
Fuel cell

An electric cell that converts the chemical energy of a fuel directly into electric energy in a continuous process. The efficiency of this conversion can be made much greater than that obtainable by thermal-power conversion. In the latter, the chemical reaction is made to produce heat by combustion. The heat is then transformed into mechanical energy by a heat engine, which drives a generator to produce electric energy. Further loss is involved if alternating current is generated and converted to direct current.

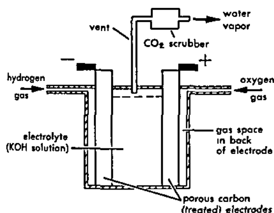
If efficiency of conversion at 25°C is 100%, then for the reaction

$$H_2(g) + \tfrac{1}{2}O_2(g) \rightarrow H_2O(l)$$

(where g and l indicate gaseous and liquid states) $\Delta H = -68.4$ kilocalories per mole of water, at 25°C, and $\Delta F = -56.6$ kcal/mole of water, at 25°C, where ΔH is the change in heat content and



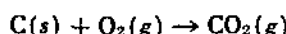
ΔF is the change in free energy. The minus sign indicates that energy is released by the reaction.



Hydrogen-oxygen fuel cell.

See ELECTRODE POTENTIAL; FREE ENERGY. The electric cell, theoretically, could yield 82.8% of the heat of the reaction in the form of electric energy.

For the reaction



$$C(s) + O_2(g) \rightarrow CO_2(g)$$

(where s indicates solid state), $\Delta H = -94.4$ kcal and $\Delta F = -94.1$ kcal. An electric cell in which this reaction occurred at 25°C could, theoretically, yield nearly 100% of the heat of reaction in the form of electric energy.

In real cells, of course, the actual output is always less than the free energy. When it is considered, however, that thermal power systems operate at not over 40% efficiencies in the largest installations, the possible gain by an electric cell process is very attractive.

Since 1894, electrochemists have attempted to make fossil fuel combine with oxygen through an electric cell process. These cells may be considered the classic fuel cells.

Principal fuel-cell reactions. The principal over-all reactions which have been employed in fuel-cell work are summarized in Tables 1 and 2.

Table 1. Theoretical cell potentials at various temperatures

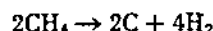
Reaction	Cell potential, volts					
	25°C	100°C	250°C	500°C	750°C	1000°C
$C + O_2 \rightarrow CO_2$	1 02	1 02	1 02	1 02	1 02	1 01
$2C + O_2 \rightarrow 2CO$	0 71	0 75	0 82	0 93	1 04	1 15
$2CO + O_2 \rightarrow 2CO_2$	1 33	1 30	1 23	1 11	1 00	0 88
$2H_2 + O_2 \rightarrow 2H_2O$	1 23	1 18	1 12	1 05	0 97	0 90

Table 2. Theoretical material consumption

Reaction	Temp., °C	Consumption, lb/kw-hr		
		Anode	Cathode	Total
$C + O_2 \rightarrow CO_2$	750	0 344	0 918	1 262
$2C + O_2 \rightarrow 2CO$	750	0 474	0 632	1 106
$2CO + O_2 \rightarrow 2CO_2$	750	1 15	0 656	1 816
$2H_2 + O_2 \rightarrow 2H_2O$	100	0 070	0 56	0 63
	750	0 085	0 68	0 763

The direct anodic use of carbon has been practically abandoned in modern fuel-cell work. Carbon potentials seem entirely due either to carbon monoxide (CO) or to hydrogen (H_2) formed at high temperature by direct reaction between the carbon and the electrolyte. For example, in the Jacques cell consisting of carbon electrodes and iron (air) electrodes in molten sodium hydroxide, H_2 is liberated at the carbon by reaction with the electrolyte. It is this H_2 which is responsible for the observed potential.

Modern fuel cells use gaseous fuels, either H_2 or CO or mixtures of these gases. The oxidizer is normally oxygen or air. Hydrocarbons have not been made to function anodically. Where potentials have been measured, they are attributed to decomposition of the hydrocarbon to liberate H_2. For example, methane decomposes at high temperatures as



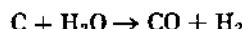
$$2CH_4 \rightarrow 2C + 4H_2$$

For technical reasons, it is simpler to use the carbon or hydrocarbon fuel in a chemical reactor to produce the active gases, H_2 and CO, than to attempt to operate a cell under the conditions best suited for the chemical reaction. Typical chemical production of the active gases might be



$$2C + O_2 \rightarrow 2CO$$

Theoretically, CO has a requirement of 1.15 lb per kilowatt-hour (lb/kw-hr) when reacted anodically against an oxygen cathode. To produce 1.15 lb CO, 0.493 lb carbon is needed. Hence a perfect process, starting with 0 493 lb carbon, would yield 1 kw-hr. A mixture of H_2 and CO can also be produced by reacting carbon with steam.



$$C + H_2O \rightarrow CO + H_2$$

The complete engineering design of the chemical reactor in conjunction with the fuel cell has not been made. Major effort, thus far, has been directed to the improvement of the fuel cell proper. The over-all efficiency of the system may be much less than that of the cell proper, however, if the reactor causes an appreciable loss of potential fuel energy.

Hydrogen-oxygen fuel cell. Work with H_2 has established that it can operate efficiently at moderate temperatures with polarization decreasing as the temperature is increased. This permits the use of aqueous solutions. One has been reported to have, at 25°C, the following characteristics:

Current density, amp/ft²	0	1	10	50
Cell voltage	1.12	1.01	0.95	0.70

In 1957 the development of a hydrogen-oxygen battery capable of long continuous duty at moderate temperatures such as 50°C was reported, but no published material has become available. It is understood to be an alkaline cell with activated carbon electrodes.

At higher temperatures, the hydrogen-oxygen battery is capable of relatively high power output. One cell using porous nickel electrodes and an alkaline electrolyte was reported to have an electrode life of some thousands of hours at a working temperature of 200°C and a pressure of 42 kg/cm² (600 psi). Characteristics of this cell are

Current density, amp/ft²	0	10	100	400	600
Cell voltage	1.10	1.02	0.895	0.73	0.60

Current efficiency was reported to be 98%. On this basis, the material consumption per kwhr has been calculated as follows:

Amp/ft²	100	400	600
Hydrogen lb/kw-hr	0.094	0.115	0.14
Oxygen, lb/kw-hr	0.754	0.924	1.12
Total	0.848	1.039	1.26

Carbon monoxide fuel cell. The use of carbon monoxide has been limited to high-temperature cells. In one carbon monoxide cell, air was used as the cathodic material. The cell voltage at 700°C was reported to fall linearly from 0.85 at no load to 0.75 at 32 amp/ft². The anodic gas was not clearly described. The cell worked with producer or illuminating gas, as well as with hydrogen. [S.EI.]

Bibliography: A. M. Adams, Chemical generation and storage of electricity, *Proc Inst Elec. Engrs.*, *(London)*, vol 2(13), 1956. A. J. Allmand and H. J T. Ellingham, *Applied Electrochemistry*, 2d ed., 1924; F T. Bacon and J S. Forrest, High-

trical Research Association, London, Ref. trans. 1B884, 1949); O. K. Davtyan, Gas cells with a solid electrolyte, *Akad. Nauk SSSR Otdel. Tekh. Nauk*, no. 2, pp. 215–218, 1946 (English trans RJ-464, Associated Technical Services).

Fuel gas

A fuel in the gaseous state whose potential heat energy can be readily *transmitted* and distributed through pipes from the point of origin directly to the place of *consumption*. The development and use of fuel gases is closely associated with the progress of civilization and mankind. Such gases have been especially prominent in the industrial development that has taken place since the eighteenth century.

Fuel gases usually include the following: natural gas, LP gas, oil gas, coke-oven gas, coal gas, producer gas, blue gas (water gas), carbureted blue gas, blast furnace gas, mixed gases, sewage gas, and synthesis gas. Typical analyses of the foregoing fuel gases are presented in the table. The analyses presented are on a dry basis, hence the Btu (heat) value of gases saturated with water vapor would be 1.74% lower at 60°F than the values in the table.

Most fuel gases are composed in whole or in part of the combustibles, hydrogen, carbon monoxide, methane, ethane, propane, butane, or oil vapors; and in some mixtures of the inerts, nitrogen, carbon dioxide, oxygen, or water vapor.

Natural gas. The generic term natural gas applies to gases commonly associated with petroliferous geologic formations. As ordinarily found, these gases are combustible, but nonflammable components such as carbon dioxide, nitrogen, and helium

The olefin hydrocarbons, carbon monoxide, and hydrogen are not present in American natural gases. The term "dry" natural gas indicates less than 0.1 gal of gasoline vapor per 1000 cubic feet; "wet" natural gas indicates more than 0.1 gal/1000 ft³. Sweet and sour are the terms applied to indicate the absence or presence of hydrogen sulfide (H_2S).

There is no single composition which might be termed typical natural gas. With the many transcontinental gas pipelines and their interconnections, the typical natural gas is approaching reality. Methane and ethane constitute the bulk of the combustible components, and CO_2 and nitrogen, the inerts. The heating value of natural gas served by a utility company is often 1000–1100 Btu/ft³.

Natural gas is an ideal fuel for heating because of its cleanliness, ease of transportation, high heat content, and the high flame temperature possible. See NATURAL GAS.

Typical gas analyses*

Type	Analysis, % vol								Specific gravity	Btu/ft³	
	CO_2	O_2	CO	H_2	CH_4	C_2H_6	C_3H_8 and C_4H_{10}	N_2		Gross	Net
Natural gas	0.2										
Propane					99.2			0.6	0 56	1007	906
Refinery oil gas						2 6	97.3	0.6	1.55	2558	2358
Coal gas		0.2	1.2	6.1	4.4	72.5		0 6	1.00	1650	1521
Producer gas	1.7	0 5	6.9	49.7	29 9		3.1	8.2	0.41	510	182
Carbureted water	8 0	0 1	23.2	17.7	1.0			50.0	0.86	143	133
gas	3 6	0.4	21.9	19.6	10.9	2.5	6.1	5.0	0.54	536	461
Sewage gas	30 0				66 0			4.0	0.86	670	601

* From L. Shnidman (ed.), *Gaseous Fuels*, 2d ed, A.G.A., 1954.

LP gas. This is an abbreviation for liquefied petroleum gas. It is a term applied to certain specific hydrocarbons which are gaseous under normal atmospheric conditions, but can be liquefied under moderate pressure at normal temperatures. *See* LIQUEFIED PETROLEUM GAS (LPG).

Oil gas. This includes a group of gases derived from oils by exposure of such oils to elevated temperatures. Refinery oil gases are those obtained as a by-product during the thermal processing of the oil in the oil refinery. They are used primarily for heating refinery equipment. A typical oil gas consists of saturated and unsaturated hydrocarbons with a heating value of 1300–2000 Btu/ft^3. Methane, ethane, propane, butane, ethylene, and propylene are the main constituents. Gas made as the main product during the gasification of oil by thermal treatment has a high heating value and is commonly termed high-Btu oil gas. High-Btu oil gas as a substitute for natural gas has played an important part in recent years in helping to meet peak loads on gas-distribution systems. A twin-generator oil-gas process has been developed to produce a gas of about 1000 Btu/ft^3. This gas consists primarily of methane, unsaturates, hydrogen, and nitrogen. The Institute of Gas Technology has developed a method of estimating the oil gasification yields for high-Btu oil-gas production.

Coal gases. A large variety of gases with different compositions and heating values are obtained through the destructive distillation of coal by heat out of contact with air. This process, known as carbonization, yields a solid residue with a higher percentage of carbon than the original coal. Bituminous coals (gas coals in particular) are used, resulting in gases of different compositions, depending upon the equipment, temperatures, and the process of manufacture.

The various coal gases differ but slightly from one another except that those produced in vertical retorts generally have a lower heating value and higher hydrogen content, but yield more tar and ammonia when steaming is increased. As the coal gas comes from the retort, its lesser components include tar, light oil, ammonia, and hydrogen sulfide. *See* COAL GASIFICATION.

Coke-oven gas. This type of gas is produced in a manner similar to retort coal gas, but the size of the charge, the operating temperature, and other features of the process are different, tending to produce larger amounts of by-products and a more uniform coke. The composition of the finished gas is similar to the other coal gases. *See* COKE.

Producer gas. Producer gas is made by the partial combustion of coal or coke, in which a mixture of air and steam is blown upward through a thick hot bed of coal or coke. The gas is high in nitrogen introduced by the air combustion. Its heating value is low, its specific gravity high, and the per cent of inerts high. Producer gas is odorless and contains 23–27% carbon monoxide. Producer gas is used as it comes from the generators after some preliminary purification. It was once the cheapest form of industrial gas, but has now been replaced largely by natural gas.

Blue gas (water gas). This is produced by the reaction of steam on incandescent carbon. The process is intermittent, consisting of alternate blow periods, during which fuel in the generator is heated by blowing air through it, and runs, or gas-making periods, during which gas is generated by passing steam through the incandescent fuel bed. Blue gas derives its name from the characteristic color of its flame, which is blue because of the high percentage of hydrogen and carbon monoxide. Because of the use of steam in its manufacture, it is also known as water gas. Blue gas is used primarily for mixing with other commercial gases.

Carbureted blue (water) gas. The manufacture of carbureted blue gas represents the combined production of blue gas and oil gas simultaneously in connected pieces of apparatus. A wide variety of oils are used, and their evaluation has been studied thoroughly.

Blast furnace gas. This gas is a by-product from the operation of a blast furnace used in the manufacture of pig iron. It is derived from the partial combustion of coke. The gas contains about 27% carbon monoxide and over 70% of inert gases (CO_2 and N_2), giving it the lowest heating value, less than 100 Btu/ft^3, of any of the commercially used fuel gases. It is used for the operation of gas engines, heating by-product coke ovens, steel-plant heating, steam raising, and crude heating. *See* FUEL; FUEL GAS SUPPLEMENTS; PYROLYSIS.

[L.L.S.]

Bibliography: *Gas Age*, 123(8):21, 1959; L. Shnidman (ed.), *Gaseous Fuels*, 2d ed., 1954

Fuel gas supplements

Types of gas produced by reforming hydrocarbons by thermal- or catalytic-cracking methods, or by reducing the heating value of propane by mixing with air. To be considered in this category, a heating gas must have burning characteristics similar to those of the gas it is designed to supplement or replace. In the industry this property is called interchange.

Since natural gas is now available to the greater part of the urban population in the United States, the major emphasis is on a gas to supplement this type of heating gas for peak-shaving (addition at times of high demand) and stand-by purposes. Of secondary interest is a supplement, or substitute, gas for the older carbureted water gas and coke-oven gas which were used universally before the transmission lines made natural gas available throughout the country. Although an ideal supplement fuel gas should have burning qualities identical with those of the gas to be substituted, the goal is not always realized and certain types of gas can be used only to supplement up to a certain degree without causing difficulties and service calls from the customers.

Thermal reforming. The Hasche process is noncatalytic, regenerative method with th

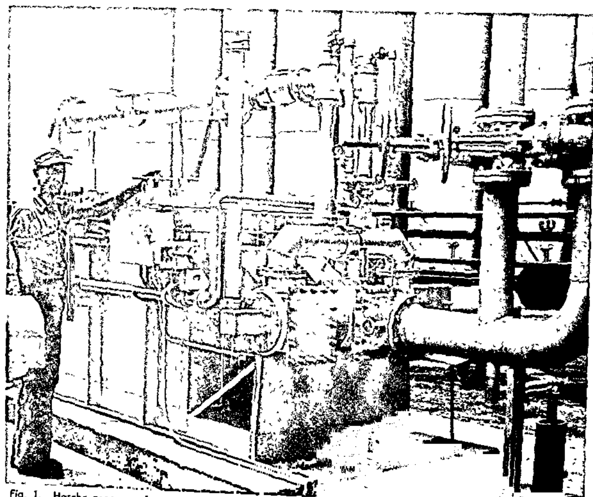


Fig. 1. Hasche process reformer.

efficiency of 95–98%. In it, a mixture of hydrocarbon gas or vapor and air is passed through a regenerative mass that is progressively hotter in the direction of the gas flow. Partial combustion occurs, liberating sufficient heat to crack the remainder of the hydrocarbons to the desired degree in a combustion zone, and the reformed gas then passes through a second regenerative mass in which it gives up its sensible heat The flow is reversed at suitable intervals to give a continuous supply of reformed gas The process is self-sustaining in heat, and the heating value of the reformed gas can be changed by varying the percentage of air in the feed The hydrocarbon charge stocks usable are propane, butane, natural gasoline, and other light distillates. No process steam is required.

One of the features of the process is the novel regenerator checker tile made of high alumina which, when assembled, gives straight, unobstructed channels of small diameter and a high ratio of mass to voids. The regenerative tiles are stacked in horizontal reformers containing two regenerative masses connected to a central combustion chamber. The reformers are of two types, straight-in-line and U-shaped. The latter is shown in Fig. 1.

The Hasche process is widely used to make both a supplement and a base-load gas of low-Btu as an interchange for complete replacement of carbureted water gas in areas where natural gas is not available. It is also used in many locations for production of high-Btu gas as a supplement and substitute for natural gas in peak shaving and stand-by. The analyses in Table 1 show gas compositions in reforming propane for supplement gas and as an interchange for carbureted water gas.

Table 1. Propane reforming product analyses, % vol

	Supplement gas	Interchange for carbureted water gas
Carbon dioxide	1 8	0 8
Acetylene	0 4	0 2
Ethylene	15 0	11.4
Propylene	3 1	6 0
Oxygen	0 6	2.1
Hydrogen	9 6	5 0
Carbon monoxide	6 4	2 8
Methane	13 5	9 5
Propane	5.1	27 7
Nitrogen	44 5	34 5
Heating value, Btu/ft^3	634	1130
Specific gravity	0 84	1 06

A recent development of the process is a modification to a semimake-and-heat operation. In countries where natural gas is unavailable, an oil gas, made from a high-carbon oil or refinery gas, is reformed to a low-Btu, low-specific-gravity gas which is an interchange for coal gas. In the United States gasoline is reformed to a gas of 1200 Btu/ft³, 0 80 specific gravity, for interchange with natural gas.

The so-called oil-gas processes are treated in another section of this article. However, there are several modifications which are adapted to produce gas supplements or interchange for natural gas which deserve particular mention. The majority of the high-Btu oil-gas units in the industry involve conversion of existing water-gas sets by installation of checkers in the generator and superheater shells. Installations include the Gas Machinery, the Hall, and the Semet-Solvay processes. Figure 2 shows a schematic diagram of the Hall oil-gas process.

Supplements also can be made by mixing oil gas with coke-oven gas and liquefied petroleum gas, LPG. There has been little need to operate these facilities because the supply of natural gas has been sufficient to avoid high-demand charges. Most of the conversions and new installations have been made in New England.

The oil-gas units operate on the make-and-heat cycle and utilize various types of hydrocarbon charging stock such as LPG, gas oils, diesel fuel, and heavy bunker grades. The latter is preferred because of its lower cost. Thermal efficiencies based on gas produced average around 60%, and if tar and light oil are included, 75–80%. In Table 2 are given analyses of oil gas obtained with the Hall process operating on various types of oils.

Catalytic reforming. There are several processes for the production of low-Btu reformed gas by the catalytic method. Among them are the UGI cyclic catalytic reforming (CCR) process, the Segas process developed in Great Britain, the French ONIA-GEGI, the Girdler, and the Surface Combustion processes. The first three mentioned are cyclic, regenerative with fixed-bed catalysts, whereas the latter two are continuous and employ externally fired tubes containing catalyst.

The CCR process consists of the conversion of carbureted water-gas sets by installing a bed of nickel catalyst in the superheater and using the carbureter as a combustion chamber and process steam superheater. It is a make-and-heat process in which fuel is burned in the combustion chamber to heat the regenerator and catalyst bed, and to remove carbon and sulfur compounds from the catalyst bed. The heat from the blast gas is recovered in a waste-heat boiler, and process steam is superheated in the regenerator at the bottom of the catalyst bed. The hydrocarbon feed is introduced between the combustion chamber and the catalyst bed.

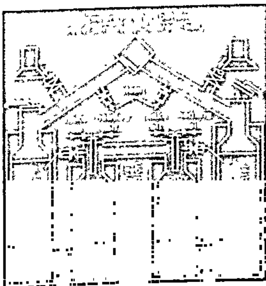


Fig. 2. Schematic diagram of Hall oil-gas process

Table 2. Analyses of oil gas by Hall process, % vol

	Conradson carbon content of oils, %				
	0 20	3 16	6 02	13 03	12 56
Carbon dioxide	4 9	5 5	4 7	6 9	6 0
Illuminants	26 2	22.4	23 4	23 6	24 3
Oxygen	0 5	0 5	0 9	0 4	0 6
Carbon monoxide	1 6	2 0	1 6	1 7	1.7
Methane	24.7	16 8	28 1	22 7	21 8
Ethane	10 5	15 8	6 6	8 9	6 8
Hydrogen	16 1	20 0	19 4	17 1	21 3
Nitrogen	15 5	17.0	15 3	18 8	17 6
Heating value, Btu/ft³	1046	1006	1047	966	974
Specific gravity	0 855	0 866	0 834	0 867	0 833

The Segas and ONIA-GEGI processes are similar to the CCR process except that the process steam and hydrocarbon flow downward through the catalyst bed The Segas catalyst is a lime-bauxite mixture bonded with bentonite.

The Girdler and Surface Combustion processes use a nickel oxide catalyst in vertical alloy tubes. The charge stock is hydrocarbon gas or vapor mixed with steam, with or without air. The tubes are externally fired to give a continuous production of reformed gas.

In Table 3 are listed typical compositions of reformed gas produced by the catalytic processes employing various hydrocarbon feed stocks.

The reformed gas made by the processes using natural gas and propane as charge stock are carrier gases and can be blended with natural gas or propane to give fuels for supplementing or interchange with blue gas, carbureted water gas, or natural gas. The reformed gases made from heavy fuel oil by the Segas and ONIA-GEGI processes can be used to supplement or replace carbureted water gas.

Table 3. Catalytic reforming processes

Feedstock analyses	CCR, natural gas	Segas, heavy fuel oil	ONIA-GEGI, heavy fuel oil	Girdler, propane	Surface combustion, propane
Illuminants	0 1	10 6	3 2		
CO_2	5 2	12.6	5 6	8 66	4 6
CO	12 5	9 9	23 6	16 48	21.1
H_2	40 4	43 6	47 2	71.81	46 3
CH_4	10 0	14 1	13 0	1 99	2 3
N_2	31.5	9 2	7.4	1 06	25.7
Heating value, Btu/ft³	266	525	411	282	230
Specific gravity	0 61	0 638	0 53	0.43	0 55

blue gas, respectively, or can be blended with propane to be used as a supplement for natural gas.

Propane-air. A widely used fuel-oil supplement for natural gas is a mixture of propane and air. Propane-air plants involve no processing but include storage tanks, vaporizers, steam-generating equipment, air compressors, injector-mixers and Btu control equipment. In operation, various ratios of propane and air are made in the 1050–1450 Btu range with corresponding specific gravities of 1.2–1.3, depending on the degree of substitution for natural gas *See* CRACKING; FUEL GAS; REFORMING (PETROLEUM REFINING).

[R.L.HA.]

Bibliography: G. M. Horner, The Segas process for the catalytic reforming of hydrocarbons, *Am. Gas Assoc. Proc*, 38:465–476, 1956; H. R. Linden and M. A. Elliott, Review of processes for production of high-B t.u. gas from liquid fuels, *Newsletter Inst. Gas Technol*, vol. 10, no. 3, May, 1958; J. J. Morgan, *Gasification of Hydrocarbons*, 1953.

Fuel injection

The forcing of fuel by a fuel pump into the cylinder of an internal combustion engine or into the air at the cylinder intake ports or near the air intake. In diesel engines, ignition timing is controlled by the time of fuel injection. In spark ignition engines, fuel injection avoids possible icing due to fuel evaporation in the metering or controlling systems of carburetors. In engines with continuous combustion, such as gas turbines and liquid-fueled rockets, fuel injection is necessary because the pumping action of piston-type engines is unavailable to draw fuel into the combustion chamber. An essential function of all fuel injection systems is the controlled metering and atomization of the liquid fuel jet by pressure in the fuel supply system.

Three methods of fuel injection have been developed for spark ignition engines: direct injection, metered port injection, and continuous flow injection. Direct injection supplies fuel to individual cylinders, either by individual metering pumps as in diesel practice or by a single metering pump with a distributor which directs each discharge stroke to an individual cylinder in sequence. In metered port injection the fuel required for a working cycle is metered to each intake port before or while the intake valve is open. This may be accomplished by a variable displacement device or control of the length of time the spray valve is opened to a constant-pressure fuel supply. Continuous flow injection supplies fuel to the intake ports, the quantity being controlled by variation of the fuel supply pressure.

Fuel injection supplies more even distribution of fuel to individual cylinders than a carburetor-and-manifold combination, and because little or no heat is necessary for vaporization of fuel prior to its distribution, more power may be developed with less tendency to knock or to vapor lock.

In a diesel engine, a fuel pump starts injection at the proper engine crank angle and meters the required quantity of fuel through the nozzle. Commonly a cam-driven plunger in a closely fitting barrel forces the fuel past a check valve, through a small bore tube, to the nozzle. Injection timing is controlled by adjusting the position of the pump camshaft relative to the engine crankshaft. Fuel quantity is controlled by adjusting the position of an overflow port that is opened by the pump plunger, the adjustment being made manually by the operator or automatically by a governor. *See* DIESEL ENGINE; GOVERNOR; NOZZLE.

In a spark ignition engine, fuel injection replaces a conventional venturi carburetor. Intermediate between full fuel injection and carburetor induction is the injection carburetor used on piston aircraft engines, in which a fuel pump forces fuel into the air stream at the supercharger inlet. *See* CARBURETOR.

In turbines and related engines, one or two pumps pressurize the fuel, although pressure in the line to the fuel nozzles is usually regulated by a relief valve. Fuel flows continuously from the nozzles and is atomized and mixed with the compressed air entering through the liner of the combustion chamber. *See* TURBINE PROPULSION; TURBOJET; TURBOPROP.

In liquid-fueled rockets, a turbine-type pump forces fuel and oxidizers through orifices in the injector assembly. *See* ROCKET ENGINE.

[N.M.]

Bibliography: J. Dolza et al., General Motors fuel injection system, *SAE Trans.*, p. 739, 1957; A. H. Winkler et al., Electrojector—Bendix electronic fuel injection system, *SAE Trans.*, p. 758, 1957.

Fuel oil

While coal and lignite tar oils are sometimes used as fuels, the usual fuel oil is a petroleum product less volatile than gasoline. For a small automatic househeating furnace, where cost is less important than reliability, something similar to kerosine is required. For heating an open-hearth furnace in steel-making, where fuel consumption is large and under close technical control, and where cost is important, a heavy unrefined, preferably aromatic, oil is used. This gives a slow-burning yellow flame

that supplies a large proportion of radiant heat.

The petroleum materials used as fuels fall into several classes:

1. *Certain unusual crude oils which are low in* gasoline and kerosine content and usually high in sulfur and asphalt. Some Mexican and Far Eastern crudes have at times fallen in this class.

2. Crude oil residua. These supply most of the heavy industrial and bunker (marine) fuel oils. They remain following the distilling off of the easily volatile gasoline, kerosine, and gas oil stocks. The residua from low-grade crudes thus contain most of the sulfur, wax, and asphalt originally in the crude; they may be quite viscous and are easily handled only when hot.

3. Distilled fuel oils, gas oils. The name gas oil was applied originally because these materials were employed for the carbureting or enriching of water gas. The gas oils are the fractions of a normal crude boiling between kerosine and the lubricating stocks. For a lower-grade petroleum they represent anything which can be distilled off (after kerosine) without drastic cracking. A conventional gas oil may boil between 450 and 750°F and have a specific gravity of 0.80–0.86. They are rather high-priced because of their preferred use as nonviscous fuels and because of competitive demand inside the petrol...

tiuid),

All the above materials may be blended to meet the tests and specifications applied. The principal tests are flash point (should be above 150°F), to indicate danger of explosion in handling; pour point, to indicate solidifying temperature; water and sediment, to control foaming and clogging, sulfur content, to control objectionable sulfur dioxide formed on burning; carbon residue because of its relation to coking; ash content; and viscosity.

The chief technical advantages of fuel oils over other fuels are greater heating value of a given weight or volume of oil (less storage space is required); greater efficiency of burning because of better contact with air; greater ease of handling and fire tending, with virtual absence of ash, and greater heat input from a given combustion space.

Oil burners have been developed to use efficiently

t

a

c

kerosine and light gas oil are employed for these services in vaporizing burners, which convert the fuel into vapor by contact with a hot surface, and in mechanical atomizing burners. Combining either of these burners with a forced draft makes it possible to burn a wider variety of oils. For large-scale industrial use, where cost is important and technical control adequate, heavy gas oils, residua from distillation, and cracked tars are burned by atomizing with air or steam. Very viscous oils can be burned by heating and circulating them past the burners; only a portion is withdrawn from the circulating stream.

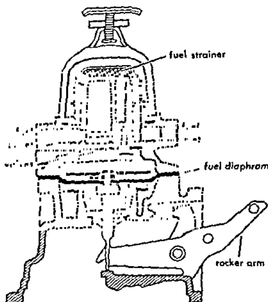
Ash content and composition have a significant effect on the equipment in which the oil is burned. Ash constituents, such as vanadium oxide, may have a highly corrosive action on the refractories which line fire boxes and on the blades of oil-fired gas turbines; similarly, the *gravity of the oil* (usually expressed in degrees API, or degrees Baumé) is important because heavier oils show greater heating value per gallon but less per pound. (Specifications for the ordinary range of fuel oils are given by the American Society for Testing Materials in Standards on Petroleum Products, published annually.) *See* OIL ANALYSIS; OIL FURNACE; PETROLEUM PRODUCTS.

[M.SO.]

Bibliography: R. T. Haslam and R. P. Russell, *Fuels and Their Combustion*, 1926; H. A. Romp, *Oil Burning*, 1937.

Fuel pump

A liquid pump for drawing fuel from a storage tank and forcing it to an engine or furnace combustor. In systems in which fuel is sprayed from a nozzle, the pump develops several hundred pounds per



Automobile fuel pump. (A.C. Spark Plug Div., General Motors Corp.)

square inch of pressure to produce atomization. Such pumps may be of the reciprocating plunger type, which can time injection for ignition timing, rotary gear, or turbine. In low-pressure fuel systems used with carburetors as in automobiles, the pump is a reciprocating diaphragm as illustrated. *See* FUEL INJECTION.

[F.H.R.]

Fuel system

A fuel system stores fuel for present use and delivers it as needed to an engine. The commonly used components for automobiles and stationary gasoline engines are fuel tank, remote reading fuel gage, filter, fuel pump (usually engine driven), and carburetor (Fig. 1)

Automotive fuel system. In automobile practice the fuel pump is usually of the diaphragm type (Fig. 2). The rotation of an eccentric on the engine cam shaft actuates a lever that pulls the pump diaphragm down against the pressure of the diaphragm spring. This lowers the pressure in the pumping chamber, drawing fuel up from the tank, through the inlet valve, and into the chamber. As the eccentric moves away from the lever, the diaphragm is moved upward by the diaphragm spring, and fuel is forced through the outlet valve to the carburetor. When the carburetor fuel chamber is filled, its float valve closes and a back pressure is created in the fuel pump chamber, holding the diaphragm down against the pressure of the diaphragm spring. The lever still rides on the eccentric but without further pumping effect. *See* FUEL PUMP.

Because the pump is usually located higher than the tank level, there may be some tendency, particularly during short engine stops on hot days, for the volatile fuel to boil in the supply line and pumping chamber to such an extent that the full travel of the diaphragm will not raise enough pressure in the vapor content to expel it onward to the carburetor. The system is then said to be vapor locked; the usual remedy is to open the engine hood until the fuel system cools off.

One way of avoiding vapor lock is to mount a remotely driven fuel pump close to and at the bottom level of the fuel tank (as shown by the dotted outline on Fig. 1) so that there is no additional boiling between the tank and pump Such pumps usually have electric drive units which operate plungers, diaphragms, or metallic bellows pumping elements, and are designed so that their speed of operation and current consumption depend upon the rate of fuel use.

Aircraft fuel system. The presence of multiple engines and multiple tanks complicates the aircraft fuel system. Also the reduction of pressure at altitude necessitates the regular use of boost pumps, submerged in the fuel tanks, which are usually of the centrifugal type and electrically driven. These supplement the engine-driven fuel supply pumps, which are usually of the gear or eccentric vane type.

Components of a typical aircraft fuel system include one main and two auxiliary tanks with their gauges; booster, transfer, and engine-driven pumps; various selector valves; and a fuel jettisoning or defuel valve and connection, which is typical also of what would be needed for either single-point ground or flight refueling (Fig. 3). The arrangement is usually such that all the fuel supply will pass to the engines via the main tank, which is refilled as necessary from the auxiliary tanks. In case of emergency, the system selector valve may connect the auxiliary tanks to the engines directly. Tank vents, not shown, are arranged so that overflow will go safely overboard.

Because of the large quantities of fuel used, aircraft fuel systems are often contaminated with dirt, metal chips, and lint; adequate filtration is therefore essential to reliable service. *See* CARBURETOR.

[F.C.MO.]

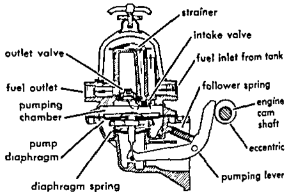


Fig. 2. Typical automobile fuel supply pump.

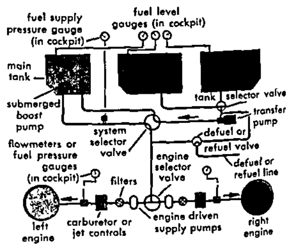


Fig. 3. Schematic of a typical aircraft fuel system.

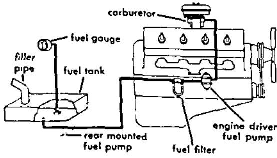


Fig 1. Elements of an automobile fuel system.

Fugacity

A function introduced by G. N. Lewis to facilitate the application of thermodynamics to real systems. Thus, when fugacities are substituted for partial pressures in the mass action equilibrium constant expression, which applies strictly only to the ideal case, a true equilibrium constant results for real systems as well.

The fugacity of a constituent i of a thermodynamic system is defined by the relation

$$\mu_i = \mu_i^* + RT \ln f_i$$

where μ_i is the chemical potential and μ_i^* is a function of temperature only, in combination with the requirement that the fugacity approach the partial pressure as the total pressure of the gas phase approaches zero. At a given temperature, this is possible only for a particular value for μ_i^*, which may be shown to correspond to the chemical potential the constituent would have as the pure gas in the ideal gas state at 1 atm pressure. This definition makes the fugacity identical to the partial pressure in the ideal gas case. For real gases, the ratio of fugacity to partial pressure, called the fugacity coefficient, will be close to unity for moderate temperatures and pressures. At low temperatures and appropriate pressures, it may be as small as 0.2 or less, whereas at sufficiently high pressures at any temperature, it can become very large.

The fugacity concept is not restricted to gaseous systems, however. Because of its relation to the chemical potential, the basic thermodynamic criterion of equilibrium requires that the fugacity of a constituent have the same value at equilibrium for every phase in which it is present. This permits the indirect determination of the fugacity for a condensed phase through the calculation of the value for the equilibrium vapor phase, for which the fugacity may be computed routinely if the dependence of its volume on temperature, pressure, and composition is known. Thus, results are readily obtained for a pure gas, but because of the more extensive data required, accurate calculations have been made for very few mixtures.

For an ideal solution, the fugacity is given by the mole fraction of the constituent times the fugacity of the pure constituent at the temperature and pressure of the solution. For liquid solutions, this is the thermodynamic counterpart of Raoult's law, but the relation applies also to ideal gaseous solutions and can serve for the prediction of the properties of real gas mixtures. Where no equation of state data are available for pure gases, their fugacities can be estimated by means of the generalized fugacity coefficient chart. At sufficiently high dilution in liquid solution, the fugacity of a nondissociating solute will become proportional to its concentration, the proportionality constant depending on the concentration scale used; this is the thermodynamic statement of Henry's law.

The dependence of the fugacity on temperature at constant pressure and composition is given by

$$\left(\frac{\partial \ln f_i}{\partial T}\right)_{P,\mathrm{comp}} = \frac{H_i^* - \bar{H}_i}{RT^2}$$

Here H_i^* is the enthalpy per mole for the constituent in the gas phase at very low pressure, and $\bar{H}_i$ its contribution per mole to the enthalpy of the system for the state under consideration. *See* ACTIVITY (THERMODYNAMICS); EQUILIBRIUM, CHEMICAL, EQUILIBRIUM, PHASE; GAS; SOLUTION; THERMODYNAMICS (CHEMICAL). [P.B.]

Bibliography: O. A. Hougen, K. M. Watson, and R. A. Ragatz, *Chemical Process Principles*, pt. 2, 1947; G. N. Lewis and M. Randall, *Thermodynamics and the Free Energy of Chemical Substances*, 1923.

Fuller's earth

Any natural earthy material (such as clay materials) which will decolorize mineral or vegetable oils to a sufficient extent to be of economic importance. It has no mineralogic significance. *See* CLAY; CLAY, COMMERCIAL.

In ancient times raw wool was cleaned by kneading the wool in water with certain earths which adsorbed the dirt and oil. The process was known as fulling; the earth, as fuller's earth. The earth used by fullers was found to have oil-decolorizing properties.

Composition. Fuller's earths are not always composed entirely of clay. Some materials which have been used for decolorizing are primarily silts with a relatively low clay content. The decolorizing power of such earths is relatively low, and they are no longer used. The clay minerals present in fuller's earth may include montmorillonite, attapulgite, and kaolinite. However, some materials containing these clay minerals often have very low decolorizing power. *See* CLAY MINERALS.

Some montmorillonite clays with low decolorizing activity in their natural state develop very high activity following treatment with acids. Such clays are not usually defined as fuller's earths. Fuller's earths possess natural activity and do not ordinarily respond satisfactorily to acid activation treatment.

Fuller's earth from different parts of the country, or even from different parts of the same deposit, frequently shows a wide variation in properties. The fuller's earth type clay is often characterized by a high water content. In a dried state it often adheres strongly to the tongue. The chemical composition of fuller's earths is extremely variable depending on its mineral composition.

Bleaching properties. The action of fuller's earth and other bleaching clays in decolorizing oils is not completely understood. This action involves a selective adsorption of color bodies and other impurities from oils. These color bodies and impurities are strongly held within the clay structure after adsorption. Because they can be removed from the

clay only by drastic treatment, it is evident that bleaching clays operate mainly by chemical adsorption.

Large tonnages of fuller's earth are used to decolorize petroleum products, cottonseed oil, tallow, soy oils, and other products. For satisfactory use the earth must only decolorize—it must not change the character of the oil or impart an undesirable taste or odor to it. Also it must not retain too much oil which cannot be reclaimed, and in some cases it must be regenerated easily by solvent or heat-treatment.

Some fuller's earths are used for other purposes, for example, oil-well drilling muds, insecticide carriers, and fillers.

Fuller's earth deposits are distributed throughout the world, but the largest deposits have been found in the United States, England, and Japan. In the United States, principal deposits are found in Georgia, Florida, Illinois, Texas, Nevada, and California [F.M.W.; R.E.GR.]

Fulminate

A compound containing the —ONC group and derived from fulminic acid, HONC. Fulminates are isomeric with cyanates; that is, cyanates have the same atoms but in different arrangement, —OCN.

The fulminates are commercially important because of the use of mercury fulminate, $Hg(ONC)_2$, in priming compositions and initial detonating agents. Mercury fulminate is very sensitive to impact, friction, and heat, but lead azide is replacing mercury fulminate as a detonating agent. *See* AZIDE; CYANATE; EXPLOSION AND EXPLOSIVE.

[E.E.WR.]

Fumagillin

An antibiotic produced by a strain of *Aspergillus fumigatus* In the commercial process, a seed fermentor, containing a medium made up of corn steep liquor, glucose, calcium carbonate, and water, is inoculated with spores of *A. fumigatus*. The seed fermentor culture is agitated and aerated for 24 hours at 26°C and then added to a production fermentor. The production fermentor medium consists of corn steep liquor, corn starch, glucose, sodium nitrate, dibasic potassium phosphate, magnesium sulfate, sodium chloride, ferrous sulfate, calcium carbonate, and water. The culture is agitated and aerated at 26°C for 112 hours and then harvested.

Chemically, fumagillin is a weak acid, having a molecular weight of 458. Its empirical formula is $C_{26}H_{34}O_7$. The structural formula is not completely known. It can be prepared in the crystalline form, which has an indefinite melting point around 192°C. Fumagillin is a mono ester of decatetraenedoic acid, $HOOC(CH{=}CH)_4COOH$, and a neutral substance, $C_{16}H_{26}O_4$. The neutral substance contains a secondary hydroxyl group, a methoxyl group, and two noncarbonyl oxygen atoms. Fumagillin is insoluble in water and saturated hydrocarbons, and soluble in most other organic solvents. It is insoluble in aqueous dilute acid and soluble in aqueous alkaline solutions Fumagillin is rather unstable and decomposes rapidly in alkaline solutions and solutions exposed to light.

Fumagillin is active against the ameba, *Endamoeba histolytica*, and exhibits activity against a sporozoan parasite, *Nosema apis*, of the honeybee, and species of *Trichomonas* and *Giardia* (flagellated protozoans). It inhibits the growth of staphylococcal bacteriophage but is without significant activity against animal viruses, bacteria, or fungi. Concentrations of fumagillin as low as 0.005–0.1 micrograms per milliliter (μg/ml) show inhibitory effects in vitro against *E. histolytica*. This antibiotic exhibits amebastatic properties at concentrations around 0.01 μg/ml and amebicidal properties at concentrations around 1.0 μg/ml.

The toxicity of fumagillin in most species of animals is relatively low. The LD_{50} in mice, the dose which kills 50% of the animals tested, is 800 milligrams per kilogram (mg/kg) subcutaneously and 2000 mg/kg orally. Daily doses of 10 mg/kg for 60 days are well tolerated by rabbits and monkeys. The toxicity with oral administration is considerably less than with parenteral administration. At high dosages, toxic manifestations in animals include anorexia, vomiting, diarrhea, intestinal hemorrhage, and weight loss.

Fumagillin is clinically effective in the treatment of human amebiasis and was used for this purpose for a short time (*see* AMEBIASIS). Oral doses ranging from 5–60 mg per day give a satisfactory clinical response. Side reactions are occasionally encountered. These include nausea, anorexia, vomiting, diarrhea, and skin rashes. The therapeutic effectiveness of fumagillin is not greatly superior to that of other amebiasis drugs and the use of this antibiotic rapidly declined because of the incidence of side reactions.

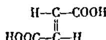
Fumagillin is effective for the treatment and control of nosema disease of honeybees, a disease caused by *Nosema apis*, a protozoan. Fumagillin is used commercially at present for this purpose. The antibiotic in the form of the dicyclohexylamine salt is added to 66% sugar syrup in concentrations of 75–100 mg of fumagillin per gallon. The bee colonies are fed 1–2 gal of medicated syrup per treatment. *See* ANTIBIOTIC. [J.C.SY.]

Bibliography: T. E. Eble and F. R. Hanson, Fumagillin, an antibiotic from *Aspergillus fumagatus* H-3, *Antibiotics and Chemotherapy*, (1):54–58, 1951; H. Katznelson and C. A. Jamieson, Control of nosema disease of honeybees with fumagillin, *Science*, 115:70–71, 1952.

Fumaric acid

An organic acid with the formula $C_4H_4O_4$. It is produced commercially by synthesis and, more recently, by fermentation. Industrial uses for this compound are, as yet, potential rather than actual, because there is no industrial need for the product.

If the microbiological production costs can be reduced, the compound may be (1) converted, by heating, to maleic acid and anhydride, which are consumed in large quantities in synthetic resins, paints, and varnishes; (2) used in printers' inks; (3) converted to glycerol-type resins for use in varnishes. The acid has a melting point of 287°C, and the following structural formula



The fermentation process produces weight yields of fumaric acid equivalent to 60–70% of sugar fermented. Selected strains of fungi of the order Phycomycetes, such as *Rhizopus nigricans*, are used (*see* PHYCOMYCETES). Cheap fermentable sugars, such as invert sugar, glucose, fructose, maltose, and starches, may be used. Untreated molasses is unsuitable. A sterilized medium, consisting of purified sugars (5–15%) and nutrient minerals, is inoculated with spores or with vegetative mycelium. Optimal temperature is 30–35°C The culture is vigorously aerated and agitated during growth. The fumaric acid is continuously neutralized by addition of sterile sodium hydroxide solution, or by initial inclusion of excess calcium carbonate, or both. Free fumaric acid crystallizes from the filtered fermentation liquor upon addition of hydrochloric acid and cooling. *See* INDUSTRIAL MICROBIOLOGY. [J.W.F.]

Fumigant

A chemical compound which acts in the gaseous state to destroy pests such as insects, nematodes, arachnids, rodents, weeds, and fungi. Fumigants may be applied as solids, liquids, or gases, but to be effective they must be readily volatile so that as a gas they may reach the organism to be destroyed. Fumigants are most useful for pest control in confined or inaccessible locations. They are used regularly to control insects and rodents in flour and feed mills, grain storage structures, greenhouses, warehouses, and the holds of ships carrying agricultural and food products. Fumigants are widely used to control nematodes, weeds, insects, and fungi in soils. They are also used to control household insects, such as clothes moths, carpet beetles, and cockroaches. Closets, rooms, and entire houses are often fumigated.

Although the quantities of fumigants used are smaller than those of the corresponding sprays and dusts used for pest control, it is probable that more than $50,000,000 is spent annually in the United States for fumigants. The value of soil fumigants used annually in this country has been estimated at about $15,000,000.

For the fumigation of agricultural crops such as grain in storage, the following chemicals are commonly used: acrylonitrile, carbon disulfide (usually combined with carbon tetrachloride to reduce the fire hazard), dichloroethyl ether, ethylene dibromide, ethylene dichloride, ethylene oxide (formulated with carbon dioxide), hydrogen cyanide, methyl bromide, and propylene dichloride.

Household fumigants include *p*-dichlorobenzene for clothes moth and carpet beetle control, *o*-dichlorobenzene for termite control, and naphthalene. Professional fumigators frequently use hydrogen cyanide or methyl bromide.

Fungi which attack plants are present in most soils, and in warm climates, nematodes and soil insects greatly reduce the yield of agricultural crops. For soil fumigation in small areas, such as greenhouses and seedbeds, formaldehyde, methyl bromide, or chloropicrin is used, usually with covers of plastic, canvas, or paper to confine the gases For soil fumigation in the field, tractor-drawn applicators have been devised to inject the fumigant (usually in liquid form) a few inches below the soil surface In this way, large areas may be fumigated rapidly and economically, using chloropicrin, 1,2-dibromo-3-chloropropane (Nemagon), dichloropropene (alone or mixed with dichloropropane as DD Mixture), 3,5-dimethyltetrahydro-1,3,5,-thiadiazine-2-thione (Mylone), or sodium *N*-methyl dithiocarbamate.

Less commonly used fumigants include calcium cyanide (Cyanogas) to produce hydrogen cyanide for insect and rodent control, and diphenyl, used to impregnate fruit wrappers.

Many insecticides used as sprays or dusts have fumigant action *See* AGRICULTURAL CHEMISTRY; FUNGISTAT AND FUNGICIDE; HERBICIDE; INSECTICIDE; MITICIDE; NEMATOCIDE; SOIL STERILIZATION. [D.E.H.F.]

Bibliography: D. E. H. Frear, *Chemistry of the Pesticides*, 3d ed., 1955.

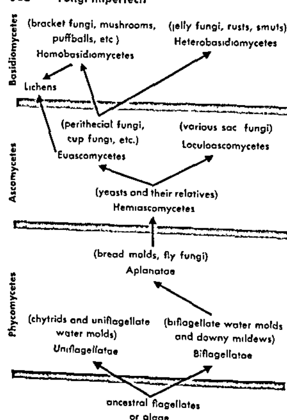
Fundamental frequency

The lowest frequency at which a system vibrates freely. The mode of vibration having this frequency is called the fundamental mode. For a complex wave consisting of a sum of sinusoidal components, that component having the lowest frequency is the fundamental and its frequency is the fundamental frequency. *See* HARMONIC (PERIODIC PHENOMENA); MODE OF VIBRATION; PARTIAL TONE. [R.W.Y.]

Fungi

Yeasts, mildews, rusts, mushrooms, and truffles are examples of fungi. Fungi are nucleated, usually filamentous, spore-bearing organisms devoid of chlorophyll; typically reproducing both sexually and asexually; living as parasites in plants, animals, or other fungi, or as saprophytes on plant or animal remains, in aquatic, marine, terrestrial, or subaerial habitats.

The so-called true fungi (Fungi, Eumycetes, Eumycophyta, Mycota) constitute a division of the plant kingdom. The division is subdivided into three classes, Phycomycetes, Ascomycetes, Basidiomycetes, and two form classes, Deuteromycetes (Fungi Imperfecti) and Lichens. In addition, some mycolo-



Diagrammatic scheme of possible relationships among major groups of fungi.

gists include the class Myxomycetes (Mycetozoa, Myxomycophyta) in the fungi. Estimates of the number of existing fungus species vary from 40,000 to more than 100,000. See ASCOMYCETES; BASIDIOMYCETES; EUMYCETES; FUNGI IMPERFECTI; LICHENS; MYXOMYCETES; PHYCOMYCETES.

Fungi, except for the yeasts, are classified almost entirely *on the basis* of their morphology and life histories, but some investigators have used physiological and chemical characteristics as well.

The most *widely* held view on the phylogeny of the fungi is based on the concept of a monophyletic origin. Accordi...

...asidiomycetes. Thus, the sporangium of the Phycomycetes is held to have given rise to the ascus which, in turn, is believed to have evolved into the basidium.

Another theory postulates a polyphyletic origin of the fungi with the lower Phycomycetes originating from flagellated *green algal* ancestors, and the Ascomycetes independently from red algal, floridean ancestors. The Phycomycetes, accordingly, have given *rise to* no other group, but the Ascomycetes have produced the Basidiomycetes. Supporters of these two views classify the fungi in the plant kingdom with *conviction*, or reluctantly for the sake of convenience.

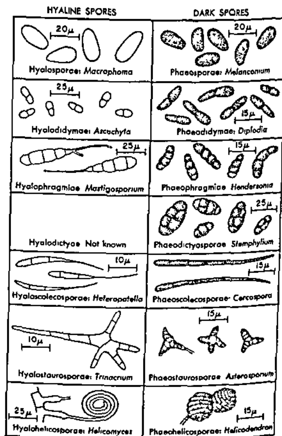
Still another view on the classification of the fungi, not widely accepted, revives and expands on E. Haeckel's idea of the Protista, and classifies the fungi in the kingdom Protoctista. *See* PROTISTA. [C.J.A.]

Bibliography: G. C. Ainsworth and G. R. Bisby, *A Dictionary of the Fungi*, 1954; C. J. Alexopoulos, *Introductory Mycology*, 1952; E. A. Bessey, *Morphology and Taxonomy of Fungi*, 1950; G. W. Martin, *Outline of the Fungi*, 1950.

Fungi Imperfecti

A class of the subdivision Fungi, or Eumycetes, of the division Thallophyta. Many members of the Fungi Imperfecti, or Deuteromycetes, are pathogenic for plants, animals, and man, causing diseases like leaf spot of maple, apple blotch and anthracnose in plants, moniliasis and chromoblastomycoses in man, and microsporoses in animals.

Fungi ordinarily reproduce by both sexual and asexual stages. L. Fuckel, a German mycologist, defined the Fungi Imperfecti as such fungi that do not have the sexual stage and thus are imperfect, that is, incomplete in their life cycle. The designations of perfect stage for the sexual stage and imperfect stage for the asexual stage are derived from this definition. The term Fungi Imperfecti is now used to designate the fungi whose sexual re-



Various types of spores and representative genera.

Spore group classification

Spore	Combined term	Spores hyaline	Spores dark
1-celled, but not filiform, or threadlike	Amerosporae	Hyalosporae	Phaeosporae
2-celled	Didymosporae	Hyalodidymae	Phaeodidymae
3- to many-celled	Phragmosporae	Hyalophragmiae	Phaeophragmiae
Muriform—has both cross and longitudinal septae*	Dictyosporae	Hyalodictyae	Phaeodictyosporae
Filiform*	Scolecosporae	Hyaloscolecosporae	Phaeoscolecosporae
Star-shaped or forked*	Staurosporae	Hyalostaurosporae	Phaeostaurosporae
Spirally coiled, septate*	Helicosporae	Hyalohelicosporae	Phaeohelicosporae

*Spores 3- to many-celled.

productive stages are unknown. The synonym, Deuteromycetes, means fungi reproducing by secondary spores. There are 1400 genera and 15,000 species known in the Fungi Imperfecti; most of these are imperfect stages of the Ascomycetes, some are Mycelia Sterilia, and a small number are Basidiomycetes. Imperfect, or conidial, stages of rusts and smuts, and Phycomycetes are generally not put in the Fungi Imperfecti.

P. A. Saccardo, the Italian mycologist, originally classified the Fungi Imperfecti into four orders: Sphaeropsidales, Melanconiales, Moniliales, and Mycelia Sterilia. In addition to the four orders proposed by Saccardo for the Fungi Imperfecti, the order Cryptococcales has been suggested for the yeasts or yeastlike organisms whose perfect or sexual stages are not known. It should be noted, however, that there are some authorities who prefer to place these organisms in the family Cryptococcaceae in the order Moniliales. A classification of the Fungi Imperfecti follows.

Order Sphaeropsidales
- Family Sphaerioidaceae
- Family Zythiaceae
- Family Leptostromataceae
- Family Discellaceae

Order Melanconiales
- Family Melancoliaceae

Order Moniliales
- Family Moniliaceae
- Family Dematiaceae
- Family Stilbellaceae
- Family Tuberculariaceae

Order Mycelia Sterilia

Order Cryptococcales

The genera, with the exception of those in the Mycelia Sterilia and Cryptococcales, are usually arranged in spore groups. The spore group classification depends on the number of cells in the spore and on its shape. Each spore group is further subdivided into groups with pale-colored, or hyaline, spores, and with dark spores.

The spores are further differentiated into slimy spores (gloiospores), adapted for dispersal by water or insects, and dry spores (xerospores), adapted for dispersal by wind. See CRYPTOCOCCALES; EUMYCETES; MELANCONIALES; MONILIALES; MYCELIA STERILIA; SPHAEROPSIDALES. [N.F.B.]

Fungistat and fungicide

Synthetic or biosynthetic compounds used to control fungal infections in man and plants.

Chemotherapy in man. The results from use of antibiotics and other drugs in the treatment of fungus infections of man have been discouraging; however, actinomycosis responds to large doses of penicillin, nocardiosis to a combination of penicillin and sulfadiazine, and paracoccidioidomycosis to sulfadiazine or sulfamerazine.

Sporotrichosis is the only mycotic disease for which a specific chemotherapeutic agent exists—potassium iodide. This compound has been used in the treatment of pulmonary candidiasis (moniliasis) with some success; it has also been used in conjunction with autogenous vaccines in pulmonary blastomycosis with less than satisfactory results.

Actidione, an antibiotic isolated from *Streptomyces griseus*, is selective for fungi not pathogenic to man. Although in vitro studies have indicated that it is active against *Cryptococcus neoformans*, a yeastlike organism causing serious infections of the central nervous system, this drug is too toxic for clinical use. Undecylenic acid, sodium propionate, and ointments containing sulfur and salicylic acid have been used in the treatment of ringworm of the smooth skin; sodium propionate has also been used with success in instances of vulvovaginitis due to species of *Candida*. Fungus infections of the hair and nails, as well as the more resistant infections of the skin caused by *Trichophyton rubrum*, have failed to respond to these drugs.

The recent development of three new antifungal agents has given new hope in the treatment of mycotic diseases of man. Nystatin, produced by *Streptomyces noursei*, is being used in the treatment of candidiasis of the skin, mucous membranes, intestinal tract, and lungs. It has little effect upon the course of disseminated candidiasis or infections caused by other fungi. Amphotericin B, also isolated from a species of *Streptomyces*, promises to be valuable in the treatment of histoplasmosis, blastomycosis, and coccidioidomycosis. Although the clinical response of patients with cryptococcal meningitis treated with amphotericin B has been disappointing, it still remains the only chemotherapeutic agent available for this disease.

Griseofulvin is an antifungal agent, produced by *Penicillium griseofulvum*, with a marked activity against those fungi causing ringworm. Although it only became available commercially in 1959, limited experience with griseofulvin indicates that ringworm of the hair or nails, as well as the sistant skin lesions due to *T. rubrum*, may un

edly be treated successfully with this drug. Few data are yet available with regard to the rate of relapse following the use of griseofulvin or the failure of lesions to respond to the drug. Despite the lack of such information, griseofulvin is the first antifungal agent that promises to bring relief to the many individuals who have ringworm. *See* ANTIBIOTIC; MYCOLOGY, MEDICAL; SULFA DRUGS.

[L.D.H.]

Agricultural fungicides. These are chemical compounds used to control plant diseases caused by fungi. In 1956, $82,000,000 worth of fungicides was sold to help prevent the $3,500,000,000 loss due to the ravages of fungi on American crops. Fungicides now used include both inorganic and organic compounds. Agricultural fungicides must have certain properties and conform to very strict regulations; and of thousands of compounds tested, few have reached the farmers' fields. Special equipment is needed to apply fungicides. *See* PLANT DISEASE.

Inorganic fungicides. Inorganic fungicides, such as bordeaux mixture and sulfur, are still used in the greatest amounts. Bordeaux mixture is made by mixing a solution of copper sulfate with a suspension of lime (calcium hydroxide). In 1956, 108,000,000 lb of copper sulfate, worth about $19,400,000, was sold to make bordeaux mixture; 150,000,000 lb of processed sulfur was sold in 1956 for use as a fungicide. This quantity of sulfur was valued at about $10,500,000.

Organic fungicides. Organic fungicides have become increasingly important since 1934. Some of the most useful are derivatives of dithiocarbamic acid. Examples are ferbam and ziram, the iron and zinc salts, respectively, of dimethyldithiocarbamic acid; and nabam, zineb, and maneb, the sodium zinc, and manganese salts, respectively, of ethylenebis(dithiocarbamic acid). Another related fungicide of importance is thiram or bis(dimethylthiocarbamoyl)sulfide. Other representative fungicides widely used are captan, *N*-(trichloro-

gly-

ranil,

2,3-

dichloro-1,4-naphthoquinone. In addition to these

crotonate. Two antibiotics, streptomycin for tobacco blue mold and cycloheximide for cherry leaf spot, are used as fungicides.

Requirements and regulations. These stipulations must be met by manufacturers before they sell fungicides. Agricultural fungicides must be registered with the U.S. Department of Agriculture. To get acceptance, manufacturers must prove that their products control stated diseases, without plant injury, when properly applied. In addition, the Miller law, administered by the Food and Drug Administration, will not allow materials to leave poisonous residues on edible crops. These requirements needed to protect the public have increased the cost of developing new agricultural fungicides Labels on agricultural fungicides must state the uses for which the product is intended and the conditions of use.

Formulation. The manner in which these compounds are applied, that is, as wettable powders, dusts, or emulsions, is often essential to the success of agricultural fungicides. Raw fungicides must be pulverized to uniform particles of the most effective size, mixed with wetting agents, or dissolved in solvents. These carriers or diluents must neither degrade the fungicides nor injure the plants.

Foliage fungicides. This type of fungicide is applied to aboveground parts of plants, usually to prevent disease rather than cure it. Because they are intended to form a protective coating on the plant surface that kills fungus spores before infection occurs, foliage fungicides must adhere to foliage despite weathering. Fungicides also must be sufficiently stable chemically to resist degradation by water, oxygen, carbon dioxide, and sunlight. Sometimes, as in the case of zineb, specific chemical changes by weathering are necessary to produce highly fungicidal derivatives. Protective fungicides must be insoluble in water in order to remain on foliage. Certain foliage fungicides, however, are water-soluble. These materials destroy the fungus in diseased spots after infection. Fungicides of this type are called eradicant or contact fungicides. An example is 2-(phenylmercuriamino)ethanol acetate.

Seed treatments and soil treatments. Seeds and seedlings are protected against fungi in the soil by treating the seed and the soil with fungicides Seed-treating materials must be safe for seeds and must resist degradation by soil and soil microorganisms. Fungicides mixed with soil may be harmful to seeds and plants and still be useful. Materials of this type, such as formaldehyde, are usable because they are volatile. Used before planting, they have a chance to kill soil fungi and then escape from the soil.

Chemotherapeutants. Chemotherapeutants are compounds which permeate the plant to protect new growth or to eliminate infections which have already occurred. At present, there is a great deal of interest in finding plant chemotherapeutants The most promising of these materials are the oxime and semicarbazone of cycloheximide.

Application methods. Dusters and sprayers are used to apply foliage fungicides. Conventional

ad-

de-

velopments in spray equipment are the mist blower and the low-pressure, low-volume sprayer. The mist blower uses an air blast to spray droplets onto foliage. Mist blowers have been successful for applying fungicides to trees, but less satisfactory for applying fungicides to row crops. The low-pressure,

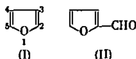
low-volume sprayers are lightweight machines which apply about 80 gal of concentrated spray liquid per acre at a pressure of about 100 psi. These have been successfully used to protect tomatoes and potatoes against diseases caused by fungi. *See* AGRICULTURAL CHEMISTRY.

[J.C.HO.; S.RI.]

Bibliography: J. G. Horsfall, *Principles of Fungicidal Action*, 1956.

Furan

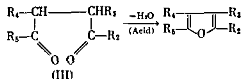
One of a group of organic heterocyclic compounds containing a diunsaturated ring of four carbon atoms and one oxygen atom. Furan (I) is a typical member of the group. *See* HETEROCYCLIC COMPOUNDS. Furfural (II) and some of its close rela-



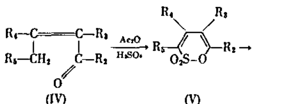
tives such as furfuryl alcohol, tetrahydrofurfuryl alcohol, and tetrahydrofuran are important chemicals of commerce. *See* FURFURAL.

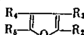
Properties and preparation. Furan (I) is a colorless, volatile liquid, bp 31.4°C, n_D^{20} 1.42140, density (20/4) 0.9378, which is stable to alkali but not to mineral acid. Its water solubility is approximately 1% at room temperature. On exposure to air, furan decomposes by autoxidation. Substituted furans, particularly negatively substituted furans, are less sensitive. The furan system is aromatic, with resonance energies of 17–25 kcal/mole reported. Nitration, halogenation, acylation, mercuration, and sulfonation reactions occur with relative ease. The extreme sensitivity of furan itself to strong acid precludes many direct electrophilic substitutions. The presence of negative groups stabilizes the ring. Unoccupied 2 or 5 positions are invariably favored as the site of substitution. Addition reactions, in contradistinction to substitution reactions, have also been observed. Furan, for example, condenses with dienophiles in the Diels-Alder reaction.

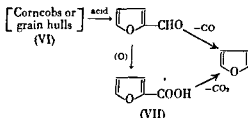
The most versatile general furan synthesis is the acid-catalyzed cyclization of 1,4-dicarbonyl compounds (III). The R groups in (III) may be hy-



drogen, aryl, alkyl, or carbethoxyl. Another synthesis (Treibs) transforms an α,β-unsaturated ketone (IV) to an unsaturated δ-sultone (V), which on pyrolysis loses sulfur dioxide and forms the furan. Furans are also formed in the acid-catalyzed dehydration of carbohydrates, the most important







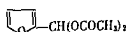
example being the production of furfural from the polysaccharides present in cheap agricultural waste materials (VI). Furan is obtained from furfural either directly by catalytic decarbonylation or by oxidation to furoic acid (VII) followed by decarboxylation.

Important derivatives. Oxidation of furans generally disrupts the system, although some ring oxidations give useful products. Ring reduction is possible, the standard method being catalytic rather than chemical. Several products derived from furfural directly or indirectly by hydrogenation are of considerable industrial importance.

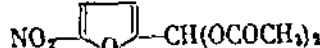
Reaction of furans with halogens, although not altogether free of complications, can be used for the synthesis of halofurans. Furansulfonic acids are prepared by sulfonation with sulfuric acid, or with a sulfur trioxide–pyridine complex. Nitration of furan compounds generally makes use of fuming nitric acid in acetic anhydride solvent. The synthesis of the bacteriostatic and bacteriocidal agent nitrofurazone, Furacin (IX), involves such a nitration of derivative (VIII). *See* NITROFURAN.

Furan aldehydes, prepared by formylation with hydrogen cyanide and hydrogen chloride, resemble benzaldehyde in their functional group reactions. Furyl ketones, prepared by Friedel-Crafts acylation, react normally. All furan mono- and polycarboxylic acids are known. Removal of the carboxyl group from the 2 or 5 positions by decarboxylation is easier than from the 3 or 4 positions.

Tetrahydrofuran (XI) is prepared either by decarbonylation of furfural to furan followed by catalytic hydrogenation, or by the Reppe process from butynediol (X). Tetrahydrofuran is a water-miscible, colorless liquid, bp 66°C, density (20/4) 0.8880, and n_D^{25} 1.4040. It is an industrial solvent, and also serves as an intermediate in the synthesis of butadiene for polymerization, adiponitrile (XII)

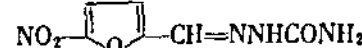


(VIII)

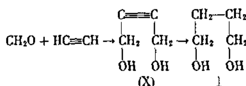


H_2O



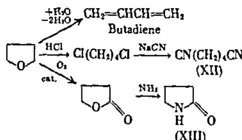


(IX)





(XI)



for nylon, and pyrrolidone (XIII) for the blood extender polyvinylpyrrolidone. *See* PYRAN.

[W.J.GE.]

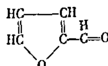
Bibliography: A. P. Dunlop and F. N. Peters, *The Furans*, Am. Chem. Soc. Monograph 119, 1953.

Furfural

When pure, a colorless liquid aldehyde boiling at 162°C and also called furfuraldehyde, fural, 2-furaldehyde, or furole. It is obtained by the digestion of corn cobs or oat hulls with dilute acid. A wide variety of agricultural by-products also yields furfural if treated in a similar manner; thus, the potential supply of this highly reactive aldehyde is virtually limitless.

Uses. Furfural was first produced on a commercial scale in 1922, and the subsequent reductions in price have aroused increasing interest in this chemical. It is used as a general synthetic intermediate in the preparation of chemicals, many of which compete with the same chemicals derived from coal and petroleum. Large amounts are also used in the preparation of molding resins and other polymers of value to the plastics industry. Its unusual solvent properties make it useful in the refining of vegetable and lubricating oils and in extracting certain components, such as butadiene, from cracked refinery gases. Other uses to which furfural has been put are as an insecticide, herbicide, fungicide, and embalming fluid.

Chemical properties. Furfural has the following structural formula:



It contains not only an aldehyde group but also an ether linkage (C—O—C) and a system of alternating single and double bonds (diene structure). These provide several sites in the molecule for reaction with other compounds. It is similar to benzaldehyde in most of its reactions as an aldehyde and in its ability to undergo substitution reactions for hydrogen atoms on the ring. Resin formation can take place by reaction at the aldehyde group or through the diene-aldehyde system (also known as a resinophore grouping). The ether linkage provides a point at which the ring can be opened to give straight-chain compounds.

Three general routes to furfural derivatives are employed in the chemical industry. It can react as an aldehyde with phenols, ketones, esters, and other materials to produce resins and compounds valuable as plasticizers. Catalytic removal of the aldehyde group from furfural gives furan, C_4H_4O, which can be hydrogenated to tetrahydrofuran. The latter is a widely used solvent and also an intermediate in the preparation of nylon and other polymers, butadiene, and numerous cyclic and straight-chain chemicals. Catalytic hydrogenation of furfural gives furfuryl alcohol and tetrahydrofurfuryl alcohol. The former has been used in the preparation of resins, and the latter can be converted to dihydropyran. Dihydropyran is either hydrogenated to give tetrahydropyran, which resembles tetrahydrofuran in its solvent properties, or converted directly to polymeric materials and chemicals important to the industry. *See* ALDEHYDE; BENZALDEHYDE; FURAN; POLYMERIZATION.

[A.E.B.]

Bibliography: A. P. Dunlap and F. N. Peters, *The Furans*, Am. Chem. Soc. Monograph 119, 1953.

Furnace (steam generating)

An enclosed space provided for the burning of fuel. The heat produced in a furnace may be used typically to generate steam for power or heat. A common example of a furnace is the domestic heating unit (Fig. 1). The space defined as th furnace ends where the gases enter the cl

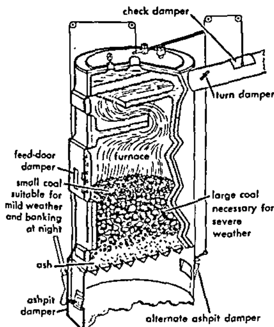


Fig. 1. Water-cooled furnace for domestic hot water boiler.

absorbing portion of the unit. Here the cold surfaces rapidly cool the gases and prevent further burning even though some combustible may still be in the gases. The furnace usually includes the fuel bed itself or other firing means. On larger units the furnace must provide sufficient residence time for the fuel to complete its combustion without severe wall impingement. Appreciable cooling of the gases in the furnace is desirable when burning low-fusing-temperature ash fuel to minimize slagging of the close-spaced heating surfaces.

Furnace walls. The boundary walls of the furnace must be practically air-tight, heat-resistant, and well braced. Refractory walls of boiler furnaces were formerly of brick, the usual wall consisting of 18 in. of firebrick backed up by 4 in. of red brick. Medium- and low-fusion coal ash rapidly washed the brick away; spalling due to temperature changes was severe; air leakage and maintenance costs were high. The gases were not cooled appreciably, resulting in troublesome slagging of convection surfaces. Later air-cooled refractory walls were used, but they also had limited life and have been superseded by water-cooled walls.

Water-cooled walls are made in a variety of patterns (Fig. 2). The effectiveness of heat absorption of the different wall tube arrangements varies with the type and also with the ratio of tube spacing to tube diameter. The membrane wall is welded airtight for use on pressure furnaces.

Hand firing. Coal is hand fired only in a few of the older small furnaces. The grates were either stationary or of the dumping type. Grates 10–12 ft long can be hand fired with anthracite, but shorter grates are necessary for bituminous coal.

Many hand-fired furnaces operated on natural draft at rates of 20–30 lb/ft² of grate per hour when burning bituminous coal. When forced draft was used, rates of 30–40 lb/ft² of grate per hour for bituminous coal were attained.

The increased cost of labor justifies the use of mechanical firing even on the smallest units. *See* STOKER.

Mechanical firing. Coal, crushed to the appropriate fineness, is delivered into the furnace by a variety of means, each adapted to feed the coal at the burning rate and to introduce air. A mechanical stoker may also remove the slag and ash; a separate grate may perform this complementary function; or the firing method may facilitate their removal.

Chain grate. The only older type of stoker still being installed has a chain grate (Fig. 3). It feeds the coal out of a hopper in a layer of uniform thickness to be burned as it is conveyed over the various air compartments. Air pressure is low in the first compartment to facilitate ignition, is at a maximum in the second and third compartments, and tapers off toward the rear. The jet ignition arrangement of air jets in the front wall stabilizes the ignition and permits the unit to respond readily to changes in load.

Traveling grates. A similar stoker has plates instead of links. Whereas the chain grate is made up of a large number of cast-iron links hinged on

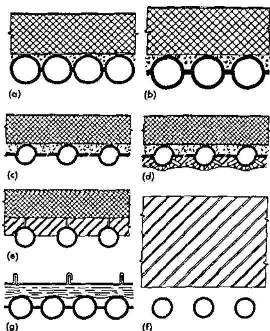


Fig. 2. Forms of water-cooled furnace walls. (a) Close-spaced tube wall. (b) Welded ligament panel wall. (c) Flat studs welded to sides of tubes. (d) Full stud wall refractory covered. (e) Tube and tile wall. (f) Tubes spaced from refractory wall. (g) Membrane wall with blanket insulation.

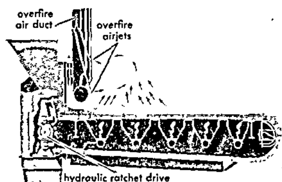


Fig 3. Chain-grate stoker with jet ignition. (Babcock and Wilcox Company)

through rods, with the air passing between the links, the traveling grate is made of cast-iron plates or grate bars fastened to several chains. The plates are perforated with air holes, usually 3/16–1/4 in in diameter.

Vibrating grate. The vibrating grate is made of water-cooled tubes covered with castings usually on 14° slope. The entire grate is vibrated to obtain the desired feed along the grate. A typical cycle is 1–11 seconds every 2 minutes. The vibrating grate is used on smaller industrial boilers.

Spreader stoker. Coal is sprinkled uniformly over the grates by a spreader, which is usually a rotating flipper. In some applications the spreader is an array of air or steam jets as on the locomotive stokers.

Traveling grates are generally used with the larger spreaders, but with smaller units dumping grates are used. Reciprocating grates may also be used. They consist of alternate overlapping grate bars that move slowly back and forth over stationary bars, thus advancing the fuel bed. Reciprocating grates have the advantage that the stroke can be reduced toward the discharge end, building up a thicker bed and giving more time to burn out the carbon. Air spaces can also be varied along the length.

With spreaders, some fines are burned in suspension, and the outer layer of all pieces is partly devolatized so that little coking takes place on the grates. Ignition is good and response to load changes is excellent. Practically all types of bituminous coal and lignites can be successfully burned. Overfire air jets help reduce smoke and unburned losses. Cinder carryover is much higher [illegible] installations. Refeeding of coarse cinders collected in rear boiler hoppers is common practice and, in some cases, the fly-ash from the collectors is refed to the furnace. This must be done with care, however, as the circulating load of cinders may build up to the point where it overloads the fly-ash collector.

Underfeed stokers. Coal can be forced into the bottom of the fuel bed by means of plungers or screws. The volatile tars pass up through the hot fuel bed where they are cracked and burned in the furnace without smoke. Fuel beds are 3–4 ft thick; only highly coking coal can be used. Large fissures due to the coking allow the air to flow through the fuel bed. Prior to the general use of pulverized coal, multiple-retort underfeed stokers were used on the largest central station boilers. Cinder carryover was excessive if burning rates were forced. Because the beds were thick, it was difficult to keep them uniform. Greater capacity is obtained per foot width with pulverized coal so that underfeed stokers are now installed only as single retorts for small industrial boilers and screw feeds for domestic units.

Pulverized coal. Hand-firing and mechanical stoking of coal are more difficult to control than is firing with gas fuel. Consequently pulverized coal suspended in air was proposed as a means of simulating gas. When coal, pulverized to a fineness of 70% through 200 mesh (0.0029 in.), is mixed with primary air, it can be blown through pipes without settling out at velocities as low as 3000 ft/min and burns much like gas. The fact that pulverized coal fired on one side of a furnace can develop 35,000,000 Btu/(hr) (ft) of width, compared with 13,000,000 Btu for stokers, led to its rapid adoption for firing large central station boilers.

Two systems of pulverizing are used. In the storage system, coal is pulverized in an air-swept pulverizer, using hot air or flue-gas for drying. The air and coal from the pulverizer are separated in a cyclone. The moist dusty air then goes to a second cyclone, a bag collector, or to the boiler furnace where the residual coal dust is burned. The pulverized coal flows by gravity, is conveyed by screw conveyor, or is transported by air to storage bins at the boiler. From the bins, the coal is fed by pocket feeders, mixed with primary air (1–1.5 lb air/lb coal), and blown to the burners as needed by the steam demand.

The advantages of this system are that dry coal is fired, which assists ignition, and pulverizers can run at full capacity during light load or off peak periods; hence less pulverizing capacity is required. The disadvantages are that a complex system of ducts is needed to obtain hot air or gas, fires may occur in pulverized coal bins, and getting rid of the dusty moist vented air is a nuisance. There is also the cost of cyclones, dust collectors, piping, fans, pulverized coal bins, feeders, and labor to operate the system.

Since 1925 the disadvantages of the storage system have led to its being almost entirely replaced by the direct-fired system, in which the pulverizer discharges coal and primary air directly to the burners. The boiler firemen operate the pulverizers, thereby lowering the labor cost. There is much less equipment to buy and maintain, no dusty air to dispose of, and no pulverized coal bins to catch fire

Hot air from the boiler air heater is used in the pulverizers without tempering when coal is very wet and is tempered with cold air if the moisture is lower. Primary air quantities vary from 1.5–2 lb/lb coal at full load and 3–6 lb/lb coal at minimum load, depending upon the method of control, type of pulverizer, and the type of coal.

Pulverizers. Three classes of pulverizers are used for coal: (1) slow-speed ball mill with classifier, (2) medium-speed ball race or bowl, and (3) high-speed such as Atrita or the impact type. The slow-speed ball mill is used for all types of coal and is especially suited to abrasive anthracite. The medium-speed pulverizers are generally used for bituminous coal and lignite. Power is low, flexibility good, and fineness is easily maintained. The high-speed pulverizers are used for bituminous coals. They have a lower first cost. *See* CRUSHING AND PULVERIZING.

Power for the pulverizer and primary air fan for 50-grindability coal and 70% through 200 mesh, varies from 13 kw-hr/ton for the lowest to 25 kw-hr/ton for the highest. At 0.7 lb coal per kw-hr, this pulverizer power is equivalent to 0.45–0.87% of the power produced.

Burners. A pulverized-coal burner consists of (1) nozzles for admitting primary air, (2) coal impellers for retarding the velocity of primary air and coal and for providing a flame holder, and (3) secondary air ports or annuli adjacent to the primary air and coal nozzles. These ports are so arranged that the secondary air mixes turbulently with the primary air and coal a short distance after both streams are in the furnace and after the coal is ignited.

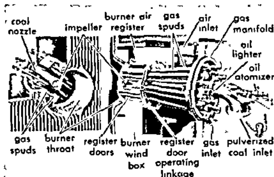


Fig. 4. Horizontal circular trifuel burner for firing pulverized coal, oil, and gas. (*Babcock and Wilcox Company*)

A horizontal circular trifuel burner for pulverized coal, oil, and gas is made in several similar types and is used on large and medium-size boilers (Fig. 4). One burner may have a capacity of 100,000,000 Btu/hr. *See* STEAM BOILER.

A tangential burner for pulverized coal, oil, or gas is used on large units (Fig. 5). In this burner

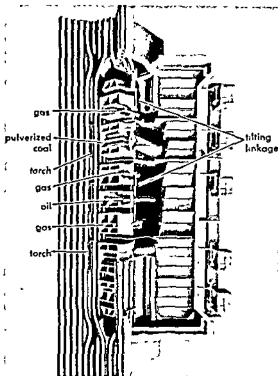


Fig 5. Tangential tilting burners for pulverized coal, oil, and gas. (*Combustion Engineering Company*)

arrangement the mixing is largely in a vortex formed in the furnace by burners, which are located in the corners of the furnace and directed tangentially to an imaginary circle in the center of the furnace.

The vertical U-flame burner arrangement developed in the early installations of pulverized coal, used vertical down-flow burners on one or both sides of the furnace. The flame path was U-shaped with secondary air coming through ports in the front wall below the burners. This arrangement was successful on bituminous coal and is still used for burning anthracite. It is not as well adapted to a combination of coal, oil, and gas as the horizontal circular burner. It is, however, well suited to burn low-volatile fuels, such as coal chars from low-temperature carbonization processes and fluid petroleum cokes.

Several burners pass the primary air and coal and secondary air between tubes on regular furnace wall spacing. The cross-tube burner, so called because it extends horizontally across the vertical furnace wall tubes, slopes 15° downward from the horizontal. It is effective on slag-tap units. Units may be installed on opposite sides of the furnace, although usually they are installed on one side only. Multitip burners, firing vertically down between the roof tubes in a down-flow furnace, are used on many dry-ash furnaces. The turbojet furnace has the opposed burners tilted downward so that they impinge on the slag-tap floor; it is used

on low-volatile petroleum fluid cokes as well as on bituminous coals.

Slag and ash. Most operating difficulties in burning coal and oil arise from the slag and ash left.

In stokers, low-fusion ash forms clinkers in the fuel bed and chokes off the air flow. Most ash in the coal goes to the stoker ash-pit, carrying with it considerable unburned carbon, representing 3–5% loss in efficiency. Some of the ash may be lifted from the fuel bed on a chain grate and, if it is low-fusion, gives troublesome deposits.

In pulverized-coal-fired furnaces, all ash is in suspension. If sufficient cooling surface is used to keep the ash essentially dry, 85–90% of the ash passes through the boiler and superheater, requiring soot blowers to keep the heating surface clean. The softening temperature of coal ash varies from 1900–2900°F and the fluid temperature from 2000–3000°F so that the gases must usually be cooled to 2100°F or lower before entering the close-spaced boiler tube banks or superheaters.

Dry-ash furnaces can be used with coals having ash softening temperatures of 2000°F and above, but portions of the furnace are hotter than the exit so that wall blowers are required to keep the water-cooled walls clean. Refractory walls will slag with almost all coals fired in pulverized form.

Coals having ash-softening temperatures below 2500°F are burned in pulverized form with wet ash removal by limiting the amount of cooling in the first part of the furnace and locating the burners so that their hot zones are close to the floor and tap hole. A two- or three-stage furnace is often used, the later stages cooling the flue gases to a temperature suitable to enter the close-spaced tube banks. The slag is drained from the first stage in molten form either continuously or intermittently into a water tank where it is quenched and disintegrated.

The ash in fuel oil is small (0.1–0.2%) but it has low-melting vanadium and sodium compounds so that tube deposits are troublesome, requiring almost as much furnace cooling as coal firing.

Cyclone furnaces. Dry-ash pulverized-coal furnaces trap 15% of the ash in the furnace and boiler; slag-tap pulverized-coal-fired units trap half the ash as molten slag. The cyclone furnace traps as much as 85% of the ash (Fig. 6). Coal crushed to ¼ in. or less is blown in tangentially at the end of the cyclone; secondary air enters through the tangential inlet at 300 ft/sec. The fines in the coal (preferably 10% or more through 200 mesh) are rapidly ignited in the turbulent cyclone, and the larger particles are thrown to the walls where they stick to the wetted slag surface. The walls are of closely spaced water tubes with full studs and chrome ore refractory to which the wetted slag adheres. The high-velocity air scrubbing the walls rapidly burns the larger particles of coal that adhere there, leaving their ash on the molten surface. This slag drains to a tap hole; the reentrant-gas outlet prevents unburned coke and slag from going out with the gases. Most coals in the northern

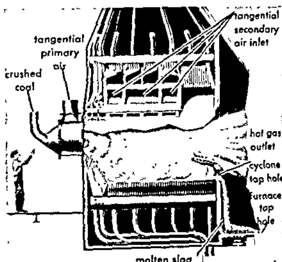


Fig. 6. Cyclone furnace for pulverized coal. (*Babcock and Wilcox Company*)

part of the United States have low enough melting temperature to work satisfactorily in a cyclone furnace.

Coal can be prepared more cheaply for a cyclone than for a pulverized-coal burner; a hammer mill using about 1.0 kw/ton is adequate, and maintenance is correspondingly less. The hot slag lining of the cyclone will reestablish ignition after the fuel has been off for as long as one minute. Oil and gas are successfully fired in cyclones as alternate fuels.

Large boilers having a steam capacity of 2,000,000 pounds per hour are fired by seven 9-ft diameter cyclones arranged on opposed sides of the furnace.

Oil and gas furnaces. For burning oil, furnaces similar to pulverized-coal furnaces are used. For oil or gas firing, however, the hopper floor is not needed; a flat or slightly sloping floor is used instead. Because there is no slag or ash in gas fuel, the cooling in the furnace can be reduced, resulting in smaller furnaces. However, most gas-fired units are designed to burn either oil or gas and sufficient cooling surface is installed to work satisfactorily with oil. *See* GAS FURNACE; OIL FURNACE.

Gas and oil are generally burned in circular burners but they are also burned in tangential burners, and in cross-tube and multitip burners and cyclones.

Oil is atomized by its own pressure or by the energy of steam or air. The Y-jet atomizer has two to six discharge holes with steam and oil mixed in each one separately. A plunger-type atomizer has a controllable plunger valve, which reduces the area of the tangential inlets to the swirl chamber, thus giving good atomization over a wide range of turndown. Return-flow atomizers maintain a high rate of oil flow through the tangential entrances to the swirl chamber at all outputs, thus obtaining good atomization. The surplus oil is returned to the oil supply system.

Typical ranges of heat available at maximum load

Furnace type	Heat available, Btu/(hr)(ft²)
Cyclone	650,000–700,000
Cyclone plus primary furnace	250,000–300,000
Cyclone plus primary and secondary furnaces	70,000–90,000
Pulverized-coal first-stage slag-tap	250,000–300,000
Pulverized-coal total slag-tap	80,000–90,000
Pulverized-coal dry-ash	80,000–90,000
Oil	100,000–140,000
Gas	140,000–175,000

Steam or air atomizers, plunger atomizers, and return-flow mechanical atomizers all have practical operating ranges of 10 to 1.

Furnace size. The size of a furnace is set by the amount of water cooling required to cool the gases and slag particles sufficiently so that the slag will not build up on close-spaced tubes of the boiler banks and superheater. Burners are so well developed that the fuel can be burned in much smaller furnaces than are usually required to provide the cooling surface.

The ratio of the total heat in the flue gases to the amount of furnace cooling surface is a measure of the reduction in temperature of the flue gases in the furnace. This ratio is called heat available and is expressed in Btu/(hr)(ft²) of cooling surface. Values for various furnaces are given in the table.

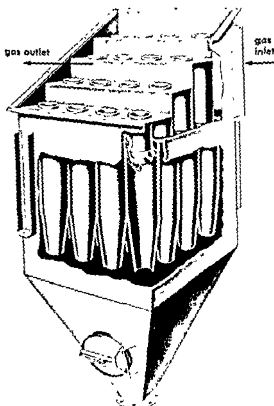


Fig. 7. Mechanical dust collector. (Western Precipitation Corporation)

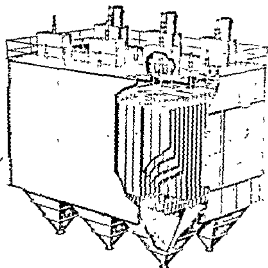


Fig. 8. Electrical dust precipitator. (Western Precipitation Corporation)

Stacks. Flue gases are removed from the plant by stacks. They are constructed of brick, concrete, or steel with or without brick lining in the lower portion. *See* CHIMNEY.

For gases at 600°F, a 200-ft stack would theoretically develop about 1½ in. water suction at the base, but actually the friction loss due to flow of gases would halve the net draft. Wind conditions affect the draft, and ratings are often limited by the weather.

Natural draft from the stack became inadequate as units became larger. Forced-draft fans force the the air through fuel beds or burners, and induced-draft fans pull the gases through the boiler at higher velocities. Now stacks are not used primarily to create draft and are either short stub stacks if the neighborhood permits, or are tall (400–500 ft) to disperse gases, SO_2, and fly-ash over as large an area as possible. Often the upper end of the stack is reduced in diameter to create a jet that will discharge the gases still farther up into the atmosphere.

Forced-draft fans are usually of the back-curved blade or multivane type running at a relatively high speed, as they handle only clean room-temperature air. *See* FAN.

Induced-draft fans are usually of the forward-curved or radial blade type and run slower because they handle gas at 250–600°F with some fly-ash. Fine fly-ash has a tendency to build up on the fan blades producing unbalance, and coarse fly-ash erodes the blades.

The speed of the induced-draft fan, or position of the fan inlet vanes or uptake damper is usually controlled to give a balanced draft (slight

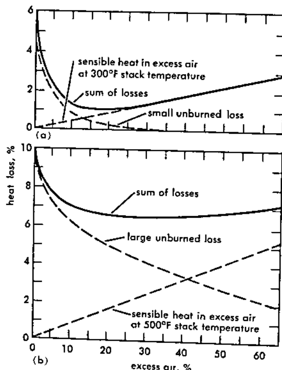


Fig. 9. Heat losses due to excess air. (a) For small amount of unburned fuel and 300°F stack temperature, and (b) for large amount of unburned fuel and 500°F stack temperature.

of 0.05–0.10 in. of water) at the top of the furnace.

Because the induced-draft fan handles hot gas at 250–600°F, the power is much higher than would be required to produce the same static pressure handling air at room temperature. Many larger units are built with a pressure-tight casing and the forced-draft fan forces the air and gas through the burners, boiler, air heater, and stack without the need of an induced draft fan. This results in a saving of about 20% in fan power, no fan maintenance, and no heat loss caused by air leakage.

Cinder trap. The cinder carry-over on underfeed stoker-fired units was so excessive that cinder traps were installed. They are usually baffles in the breeching to direct the gases down into a hopper, the cinders dropping out by inertia or against plates wetted with water. Spreader stokers also have high cinder carry-over, and hoppers at the boiler outlet are used to collect the coarse cinders, which are then refed to the furnace. The finer cinders, or fly-ash, are caught in multicyclone types of dust collectors (Fig. 7). Pulverized-coal-fired units are usually equipped with mechanical multicyclone fly-ash collectors, or electrical precipitators (Fig. 8). Both may be used in series. Many cities have ordinances limiting the dust loading to less than 0.85 lb/1000 lb flue gases. *See* SMOG.

Because of the small amount of fly-ash (8–10% of the coal ash) leaving a cyclone-fired boiler, the electrical precipitator need only be about half the

Fig 10. Pneumatic combustion control for spreader-stoker-fired boiler. (*Bailey Meter Company*)

size of the collector needed for pulverized-coal units to give the same dust loading leaving the stack.

Flue gas analysis. The largest controllable losses in the burning of fuel are in the form of sensible heat in excess air at the stack temperature or of unburned fuel resulting from inadequate air. Excess air can be accurately measured by collecting flue gas samples at a number of representative points across the breeching and analyzing them with an Orsat for carbon dioxide, oxygen, and carbon monoxide. *See* GAS ANALYSIS.

Oxygen recorders give a more representative measure of excess air with different fuels than do carbon dioxide recorders. Steam flow and air flow superimposed on the same boiler meter chart have long been the best guide for excess air and are commonly used for automatically controlling it.

Heat losses vary with excess air (Fig. 9). For example, at 300°F and 500°F stack temperature and with the assumed resulting loss due to large or small amounts of unburned fuel, the most efficient operation is obtained when the sum of these losses is a minimum. For a particular unit the losses will be somewhere between these extremes.

Automatic boiler control. The automatic control of a boiler requires control of (1) combustion air and fuel, (2) furnace draft, (3) feedwater flow, and (4) superheat and reheat temperature. *See* STEAM GENERATING UNIT.

Combustion control automatically supplies fuel and air in proportion to the load demand required to maintain a constant steam pressure. For stable control free from hunting, equilibrium pressure should decrease as load increases; for example, from 2475 psi at no load to 2400 psi at full load (see CONTROL SYSTEMS).

In the pneumatic combustion-control system the master pressure controller uses this change in steam pressure to set up a loading air pressure which increases as load increases. The relation might be typified as:

Steam pressure, psi	*Loading pressure, psi*	*% of full load fuel and air supply*
2475	5	0
2400	20	100
2375	25	133

This loading pressure is transmitted to the control drives for air and fuel. The controls readjust their positions to correspond with the loading pressure. Thus, if steam pressure drops below equilibrium for that load, the loading pressure increases and in turn increases the fuel and air supply (Fig. 10). Loading pressure controls stoker feed directly and loading pressure modified by the relation between steam flow and air flow controls the forced draft damper. The up-take damper is controlled directly by furnace draft.

On a direct-fired pulverizer, a feeder controller varies the coal feed to maintain a desired relation between primary air flow differential and pulverizer differential. With this desired amount of coal in the pulverizer, the rate of pulverized coal delivered to the burners is proportional to the primary air flow. The loading pressure thus need only control the primary-air damper to control the fuel supply.

Gas supply is usually controlled by loading pressure on the gas valve but is sometimes controlled by the ratio between steam flow and gas flow or between air flow and gas flow. This is what is known as parallel control. [R.M.M.]

Bibliography: T. Baumeister (ed.), *Marks' Mechanical Engineers' Handbook*, 6th ed., 1958.

Furnace construction

A furnace is an apparatus in which heat is liberated and transferred directly or indirectly to a solid or fluid mass for the purpose of effecting a physical or chemical change. The source of heat is the energy released in the oxidation of fossil fuel (commonly known as combustion) or the flow of electric current through adjacent semiconductors or through the mass to be heated. In recent years, scientific and engineering effort has been made to utilize nuclear and solar energy for heating purposes. Therefore, according to the source of heat and method of its application, there are four categories of furnaces: combustion, electric, nuclear, and solar, in the order of their present commercial or industrial importance. *See* ELECTRIC FURNACE.

Furnaces employing combustion vary widely in construction, depending upon the application of the heat released, whether direct or indirect. Direct heat transfer is used in regenerative refractory-type heaters, in flow systems in which reactants are injected into the combustion gases. Indirect heat transfer is employed in heaters in which the mass to be heated is kept separate from the combustion gases and made to flow in tubes which absorb and transmit the heat to the fluid to be heated. *See* FURNACE (STEAM GENERATING).

Furnaces developed for indirect heat transfer can be divided into two classes. One class of heaters is used solely for general utility purposes, such as all types of boilers, and the second class is applied in the petroleum and chemical industry as an essential unit operation in refining or processing plants. The second category of furnaces will be discussed in this article.

Directly fired furnaces are employed in oil refineries and chemical process units whenever the temperature level to which a fluid must be heated is above that attainable with utility steam. Furnaces as a heat-transfer apparatus generally cost more per thermal unit of heat transferred than conventional tubular heat exchangers. According to the kind of service, furnaces in process units should be divided into two classes:

1. Those which perform solely a heating duty, that is, raising the temperature of a fluid and effecting essentially no change of state or of chemical composition. These furnaces may be termed conventional heaters.

2. Those which handle a fluid undergoing a change during heating. The physical or chemical changes constitute an essential performance requirement. Typical applications of this order are associated with such processes as distillation or preheating of temperature-sensitive materials, pyrolysis of hydrocarbons or organic chemicals, and catalytic steam-gas reforming for the production of synthesis gas.

The conventional tubular heater came into industrial use, especially in oil refineries, about 1925 and was termed a tube still to distinguish it from the shell still, a horizontal cylindrical vessel mounted on top of a firebox. The design of these tube stills or furnaces, which is essentially the same as originally conceived by analogy with developments in steam-generating equipment, consists of pipes connected by 180° return bends forming a continuous coil and arranged in a refractory furnace setting, partly in the combustion chamber. Heat is absorbed mainly by radiation and partly in a confined flue-gas passage from the combustion chamber in which heat is absorbed mainly by convection. The fluid flow through the coil is generally countercurrent to the flow of combustion gases, first through convection tubes and then through the radiant-tube section; thus a reasonable thermal efficiency can readily be attained by providing convection tube surface to an economically justifiable extent. See TUBE-STILL HEATER.

The furnaces provided in modern process units vary considerably in outer shape, arrangement of tubes, and type and location of burners, depending mainly upon the desire of the designer or manufac-

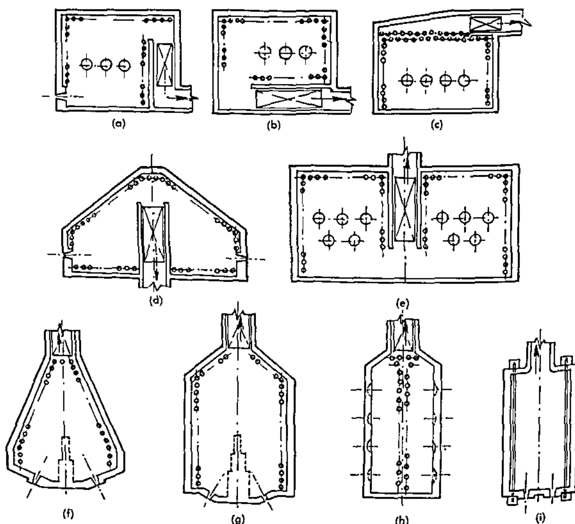


Fig. 1. Typical process furnace designs and characteristic features. (a) Radiant, up-draft, vertical-convection bank. (b) Radiant, down-draft, horizontal-convection bank. (c) Radiant [illegible] in [illegible] roc [illegible] refinery heater, dual radiant sections. (f) Up-draft furnace, slanting walls, two parallel coils. (g) Up-draft furnace, vertical walls, two parallel coils. (h) Up-draft furnace, tubes in center, burners in wall, good heat-intensity distribution. (i) Up-draft circular, all-radiant furnace, burners in floor.

turer to be identified with a specific model. Some distinctive furnace designs are shown in Fig. 1.

Design. The furnace design for a given performance or thermal efficiency is usually evolved by the following procedure: (1) determination of the composition of the combustion products and the amount of the liberated heat which must be utilized to meet the postulated thermal efficiency; (2) allocation of heat to be absorbed by the heating elements located in the combustion or radiant chamber and in the convection section; (3) determination of the heat-transfer rate and heating surface area in the radiant section; and (4) determination of the heat-transfer rate and tube surface area in the convection section or sections.

Source of heat. The combustion products vary in composition according to the type of fossil fuel burned and the excess air used in the oxidation process. The carbon and hydrocarbon content of the fuel governs the combustion-gas composition, and many formulas relating it to the elemental, oxidizable constituents of the fuel have appeared in the literature. Complete combustion is a prerequisite for high thermal efficiency, and to ensure this, air is used in excess of the minimum or stoichiometric requirements, depending upon the type of fuel and the combustion equipment or system (Table 1).

The type of fuel, its heating value, and the excess air applied in combustion determine the theoretical flame temperature which would prevail if the oxidation were instantaneous. However, combustion is a rate process requiring time, and the actual flame temperature is considerably lower because of the radiation from the combustion zone. The theoretical flame temperature can readily be determined from the heating value of the fuel and the amount of combustion products evolved, including the excess air. Accurate enthalpy data on the combustion gases are required and are presented in the literature. *See* COMBUSTION; FUEL.

In order to meet a required thermal efficiency, the heat losses must be appraised. There is a certain loss through the furnace setting which varies with the surface area of the enclosure, the heating capacity, and the atmospheric environment. The wall construction of most types of furnaces, that is,

Table 1. *Fuel-air mixtures*

Type of fuel	Type of burner	Normal range of excess air, %
Natural or refinery gas	Air and fuel gas premixed	5–15
Natural or refinery gas	Air induced by natural draft	10–25
Distillate fuel oils	Steam atomization air by natural draft	25–40
Distillate fuel oils	Mechanical atomization air by natural draft	30–50
Residual fuel oils	Steam atomization air by natural draft	40–70

Table 2. *Heat loss from setting*

Heat liberated, Btu/hr	Loss as a fraction of heat released, λ $\Delta H_{sl} = \lambda H_{fl}$
15,000,000	0.048
20,000,000	0.043
30,000,000	0.037
50,000,000	0.034
75,000,000	0.032
75,000,000+	0.030

application of insulating material, is such that the specific heat loss, Btu/(hr)(ft²) of outer wall area, is practically the same for all operating temperature levels; hence the heat loss will vary principally only with the furnace capacity. Table 2 shows the average values that will be encountered in industrial furnace constructions. The setting heat loss ΔH_{sl} can be calculated from the relation $\Delta H_{sl} = \lambda H_{fl}$, where λ is the fraction of heat released, and H_{fl} is the enthalpy of the combustion gas at the theoretical flame temperature.

The flue-gas effluent temperature from the furnace can then be determined in accordance with the postulated thermal efficiency η:

$$H_{st} = H_{fl} \times (1 - \eta) - \Delta H_{sl}$$

where H_{st} is the enthalpy of the combustion gas at the convection-bank exit temperature.

Heat load. The distribution of heat load between the radiant and convection sections of conventional heaters is associated with the manufacturer's design approach and service requirements. There are all-radiant heaters, and furnaces with tubes arranged mostly for heat absorption by convection. The amount and disposition of heating surface in the combustion or radiant chamber has a distinct influence upon the radiant-heat absorption and the supplementary heat to be recovered by convection for the required thermal efficiency. Therefore, the allocation of heating duties to the radiant and convection sections is more or less an empirical matter. Conventional process heaters are generally designed for approximately 75% thermal efficiency in the United States where fuel costs are relatively low. Most box-type heaters are designed for moderate radiant-heat-transfer rates; approximately 65–75% of the heat load is carried by the radiant section and 25–35% by the convection bank. In some cases, very high radiant-heat intensities are employed as a matter of design principle, and in these furnaces, the heat absorbed in the radiant section may be only 50% of the total duty.

Radiant section. The present procedure for determining the radiant-heat-transfer rate and heating surface rests on a long series of scientific investigations and evaluation of operating data.

The combustion gases radiate to the solid-body environment, tubes, and refractories. Carbon dioxide and water vapor are the principal constituents with emissive power. Their emissivity depends the partial pressure and the thickness of th

layer, and at 1900°F has a value of approximately 25% of the black-body radiation intensity. The terms luminous- and nonluminous-flame radiation used in scientific treatises are somewhat misleading; the mere luminosity or light has no significance as far as heat radiation is concerned, because practically all emissivity and certainly carbon dioxide (CO_2) and water (H_2O) radiation is in the infrared wavelength region, even at flame temperatures encountered in process heaters The maximum gas radiation for an infinite value of the product of the partial pressure and thickness of the gas layer, frequently termed black gas, amounts to only 21–33% of the complete black-body emissivity at temperatures of 2500 and 1100°F, respectively. The mean radiating temperature in the combustion chamber lies within this temperature range.

The consideration of heat transmission by radiation alone, from the combustion gas and refractory walls by reradiation, has not resulted in the derivation of a universally applicable formula for the heat-transfer rate in the radiant section because there is also heat transmitted to the tubes by convection which is in the order of 10% to as high as 35% of the heat absorbed in the radiant section, depending upon (1) the shape of the combustion chamber, (2) the arrangement of tubes, and (3) the type and location of burners. The convection effect is a linear function of the temperature difference (not the difference of temperatures to the fourth power) and can be varied for the same furnace design by the amount of excess air employed.

Convection heating surface. The rate of heat transfer solely by convection with flow of gas at right angle to tubes increases with the gas velocity and the temperature level, and is greater for small than for large tubes.

There are radiation effects in the convection-tube bank augmenting the heat absorption, especially when the flue-gas temperatures or the temperature differences between tube wall and flue gases are large. These include (1) radiation to the front two rows of tubes, which can be appreciable if they are exposed to the combustion chamber; (2) radiation from the hot gases surrounding the tubes; and (3) the radiation from the refractory enclosure to [illegible]

[illegible] conventional heaters, requires careful consideration of the transitory state of the fluid being heated.

Vacuum distillation of a temperature-sensitive fluid or the stripping of solvent from the raffinate and extract of a selective solvent-separation processes are typical cases in which temperature limitations are encountered in the heating operation, and partial vaporization of the fluid must take place in the tubes under relatively mild [illegible]

the proper design of the furnace coil involves exacting pressure-drop calculation for mixed vapor-liquid flow in conjunction with phase-equilibrium determination of the amount and composition of vapor and liquid at any given point in the coil. The pressure drop for mixed vapor-liquid flow is reasonably well established and vapor-liquid equilibrium relationships are known for many binary and ternary systems. The determination of temperature, pressure, and transition state of the fluid throughout the coil involves stepwise trial-and-error calculations which can best be executed by a modern digital computer for various coil dimensions and heat-intensity patterns. The coil design which satisfies all technical criteria and also cost aspects can then be selected from the computer evaluations. *See* DISTILLATION.

Gas reforming is another industrially important heating process involving a chemical change of the fluid. Methane or natural gas and steam are decomposed in the presence of excess steam in tubes filled with a catalyst. The principal products are hydrogen (H_2), carbon monoxide (CO), and CO_2, and maximum conversion of hydrocarbons is attempted. The chemical transformation proceeds at approximately 1400–1450°F and is highly endothermic, having a heat of reaction of 87,500 Btu/(lb) (mole of methane converted). The tubes are located in the center of the radiant or combustion chamber and are heated from all sides with high flue-gas turbulence around them to effect an even circumferential heat-intensity distribution. The average rate of heat transfer is moderate, in the order of 8500–9500 Btu/(ft^2) (hour), thus avoiding tube-wall temperatures greatly in excess of the catalyst and fluid-flow temperatures. *See* FUEL GAS SUPPLEMENTS.

Pyrolysis of hydrocarbons, the thermal conversion of ethane, propane, butane, and heavier hydrocarbons to produce olefins and diolefins for the petrochemical industry, is another example of a process furnace. The modern pyrolysis furnace is an outgrowth of the oil-cracking heater, the mainstay of motor-fuel production from petroleum before the advent of catalytic cracking. The design of a pyrolysis furnace must take thermochemical equilibria and reaction rates into consideration in order to realize a high conversion and a product distribution most favorable with respect to yield of olefins. The conversion of higher-molecular-weight hydrocarbons which have a high rate of decomposition involves the problem of attaining the coil outlet temperature dictated by equilibrium relationships. A shorter coil-residence time and correspondingly higher heat-transfer rates are means to this effect, but still the basic principle of moderating the heat intensity in the pyrolysis coil zone, in which the reactant is already in a high state of conversion, must be observed. Figure 2 gives a typical temperature-pressure and heat-flux pattern for a pyrolysis coil cracking a naphtha boiling in the range 190–375°F for maximum ethylene yield. The design of such a pyrolysis coil can be executed with

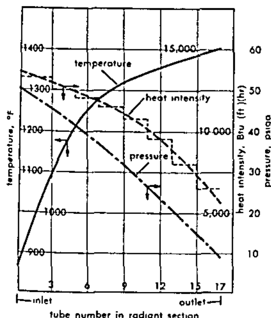


Fig. 2. Typical temperature and heat-intensity pattern, naphtha pyrolysis coil.

greatest precision or refinement by the use of a modern digital computer, permitting the rapid evaluation of all design variables. *See* CRACKING; PYROLYSIS.

Mechanical construction. The design of the furnace setting and application of refractories and insulating materials generally follow standard practice prevailing in boiler-furnace construction. Fully suspended wall construction and the use of lightweight firebrick with bulk densities of 30–65 lb/ft³ (as compared with 125–150 lb/ft³ for fireclay bricks) are generally preferred for process heaters to reduce the heat storage in the setting, [illegible]

ity burners is rather common for specific process heaters such as gas-reforming and pyrolysis furnaces.

The design of the heating coil for high-temperature service involves the application of alloy steels. Tube materials are usually selected according to the schedule in Table 3 Tubes of carbon and low-chrome steel are usually connected by forged-welding return bends, and those of stainless steel by cast $^{27}/_{12}$ chrome-nickel alloy steel bends, also weldable to the tubes.

Table 3. Heating-coil materials

Tube-wall temperature range, °F	Tube material, ASTM specification
875 and under	Carbon steel, A-161
875–1150	Low-chrome steel, A-200 (1 25–2 25% Cr, 0.5% Mo)
1150–1550	$^{18}/_{8}$ stainless steel, A-271, type 304
1550–1800	$^{25}/_{20}$ stainless steel, A-271, type 310

The flue-gas exhaust is normally carried to the atmosphere by a stack of sufficient height to induce a draft equivalent to the pressure drop in the flue-gas passage. For highest heat economy a process furnace can readily be equipped with an air preheater or waste-heat boiler usually requiring air and flue-gas exhaust blowers. Preheated combustion air results in higher flame and mean effective radiation temperatures which aid in attaining high radiant-heat-transfer rates required in certain pyrolysis operations. These furnace accessories are provided in the same manner as in boiler design practice. *See* CHIMNEY; GAS FURNACE; HEAT EXCHANGER; OIL FURNACE; PYROMETALLURGY.

[H.C.S.]

Bibliography: J. Griswold, *Fuels, Combustion and Furnaces*, 1946; A. J. Johnson and G. H. Auth, *Fuels and Combustion Handbook*, 1951; J. H. Perry, *Chemical Engineers' Handbook*, 3d ed., 1950; W. Trinks, *Industrial Furnaces*, vol. 2, 3d ed., 1955.

Fuse, electric

An expendable device for opening an electric circuit when the current therein becomes excessive. It consists principally of a section of conductor, known as a fusible element, of such properties and proportions that excessive current will melt it and thereby sever the circuit. Fuses are used in nearly all types of electric circuits to protect circuit conductors and apparatus from damage which would result from sustained excessive current.

Fuses are rated according to the voltage of the circuit for which they are designed, the current they can carry continuously, and the amount of excessive current they can successfully interrupt. Voltage ratings range from a few to over 100,000 volts, current ratings from a fraction to several thousand amperes. Interrupting ratings at lower voltages may be hundreds of thousand amperes.

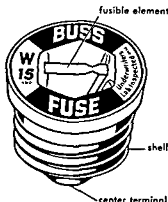


Fig. 1. Common plug-type fuse. (Bussman Mfg. Division, McGraw Edison Co.)

General classification. The most familiar type of fuse is the screw-plug fuse used in domestic electric systems (Fig. 1). It consists of a shell, similar to the base of an incandescent lamp; a fusible element connected between the screw shell and the center terminal; and a transparent protective cover.

Other fuses are cartridge fuses in which the fusible element is connected between metal ferrules at either end of an insulating tube (Fig. 2). This tube is usually filled with an insulating material, such as chalk, sand, or a suitable liquid, which cools and quenches the arc that forms when the fusible element melts.

In some types the cartridge is unfilled and open at one end to allow the arc gases to escape to the open atmosphere. These are known as expulsion fuses. Some expulsion fuses and liquid-filled fuses have a spring arranged to pull the terminals of the fusible element apart quickly when it melts, thus rapidly lengthening the arc and expediting its extinction.

There are many ratings and types of cartridge fuses ranging in size from those about 1 in. long, such as those used in automobile circuits, to high-voltage power fuses, some of which are several feet long and several inches in diameter.

Less commonly used are open link fuses, which are merely strips of fuse material bolted to open terminal blocks.

Operation. The time required for excessive current to melt a fuse varies inversely as the value of the current. However, the nature of this variation, namely the shape of the time-current curve, depends to a marked degree upon the size, shape, and material of the fusible element. The time required to quench the arc after melting depends upon the nature of the arc-quenching material as well as the fusible element. The sum of the melting and arcing times is known as the interrupting time.

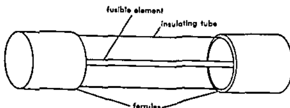


Fig. 2. Cartridge-type fuse.

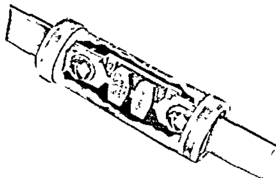


Fig. 3. Renewable-type cartridge fuse, rated 200 amp, 250 volt. (Bussman Mfg. Division, McGraw Edison Co.)

Special types. When the excess current is high, as when a short circuit occurs on a large power system, the interrupting time of some fuses is so short that the current does not have time to reach its prospective maximum value. Such fuses are known as current-limiting fuses, because they limit the amount of current that actually flows to something substantially less than that which would flow if the interrupting time were longer.

Fuses are frequently used in circuits where it is required that moderately excessive currents, such as during motor starting, be permitted to flow periodically without melting the fuse. Such fuses, called time-delay or time-lag fuses, have relatively longer melting time at moderate overcurrent than those used in circuits where periodic overcurrents do not normally occur.

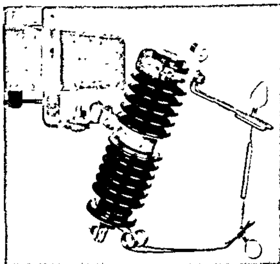


Fig. 4. Flip-open cutout fuse rated 7.8 kv, 50 amp. (General Electric Co.)

Some fuses are so constructed that a new fusible element can be inserted to replace one which has been melted. These are known as renewable fuses; an example is shown in Fig. 3.

Many fuses are equipped with a device that gives visual evidence that the fusible element has been melted. These are called indicating fuses.

Fuses used on utility-line poles are sometimes mounted in such a way that when the fuse melts, the fuse holder drops to an open circuit position readily distinguishable from the ground. These are called drop-out fuses. Figure 4 shows a type that springs open when the fuse melts, giving visual indication and providing rapid arc extinction.

A major limitation to the use of fuses arises from the fact that when a polyphase circuit is equipped with fuses a fault will usually cause only one fuse

to interrupt its phase circuit. The other phase circuits are still intact, so single-phase power is applied to anything connected to these lines. This can cause damage to polyphase motors left running single phase. This problem is overcome in some applications by an arrangement in which an open-circuited fuse will trip a multipole automatic switch or circuit breaker to open all the phase circuits. In this combination the current-limiting type of fuse is often used to interrupt only those currents which exceed the interrupting rating of the circuit breaker and occur only infrequently, leaving interruption of the more frequently occurring lower values to the circuit breaker.

For other means of protecting against excessive currents *see* CIRCUIT BREAKER; ELECTRIC PROTECTIVE DEVICES. [W.N.C.]

Fuse, explosive

A conduit that leads fire from one place to another. The firecracker fuse is a familiar example; it consists of a thin train of black powder wrapped in a tube of tissue paper. The miner's safety fuse is similar, the tissue paper being replaced by a woven and waxed fabric.

The burning rate of a black-powder fuse varies widely and depends on the grain size and physical distribution of the powder and especially on the cord construction. In pyrotechnics, a fuse called black match is made by covering a cotton cord with a coating of fine gunpowder. This is relatively slow burning, but the slight confinement afforded by a tissue paper wrapper transforms this fuse into a very fast-burning cord called quick match.

An increasingly important fuse, which is used with high explosives in mining and quarrying, is commonly known by the trade name Primacord. This material is used to initiate high explosives. A woven fabric tube of about ⅛-in. inside diameter, often reinforced with wrappings of fiber or metal, and impregnated with asphalt and wax, is filled with a core of high explosive such as pentaerythritol tetranitrate, PETN. This cord burns at the characteristic detonation rate of the explosive, about 20,000 ft/sec. Safety fuse, in contrast, burns at only about 1 ft/min.

Detonating fuse has an increasingly wide range of use. It is used to program the firing of a series of blast holes at short intervals of time. This procedure increases the efficiency and reduces the noise from blasting. A detonating cord with a thin lead sheath is also used as a precise time-delay unit in seismic work.

The word fuse should not be confused with fuze, which applies to mechanisms used to fire a shell at the moment it strikes a target, for example, or a hand grenade at a certain interval after it is thrown. Fuzes are usually mechanical or electrical mechanisms, although they may indeed involve a fuse train of black powder (a circumstance that adds to the confusion). *See* DETONATOR; EXPLOSION AND EXPLOSIVE; PYROTECHNICS. [W.E.C.]

Fused-salt phase equilibria

Much of the recent work on phase diagrams involving molten salts has been carried out as a consequence of industrial applications for liquids which are stable at high temperatures. In most cases, low freezing points and low vapor pressures are primary considerations. To some extent these requirements are conflicting, because salts with an appreciable covalent character which are useful in obtaining low freezing points also tend to give increased vapor pressure. This discussion is confined to condensed systems which are sufficiently saltlike to provide ionic melts of low vapor pressure as well as good thermal stability. According to V. M. Goldschmidt's interpretation, saltlike materials were derived from components such as H_2O, CO_2, SO_3, Cl_2, and F_2 which were volatilized from molten magma during its crystallization. Crystallization equilibria in fused-salt systems, however, provide conveniently studied demonstrations of the mechanisms occurring in the formation of igneous rocks, and were of considerable significance in establishing the foundations for phase studies in the period following the enunciation of the phase rule by J. W. Gibbs. *See* EQUILIBRIUM, PHASE.

A close connection between fused-salt phase diagrams and geochemistry stems from the model principle initially developed by Goldschmidt, who noted that isomorphic structures are assumed by ions of the same proportionate size and stoichiometric relations, but of different charge. Thus the fluorides of beryllium, calcium, and magnesium, for example, are structural models for SiO_2, TiO_2, and ZrO_2. The fluoride structures are referred to as weakened models because of the smaller electrostatic forces resulting from smaller ionic charges; they have been useful for comparisons with oxide and silicate systems.

Fused-salt mixtures were of great interest before 1900 as media for the electrolytic reduction of cations to metals. The bulk of research on fused-salt phase equilibria has stemmed from interest in electrometallurgical processes from that time until recent years when it was shown that fused-salt mixtures may serve successfully as nuclear reactor fuels, fuel solvents, coolant fluids, moderators, nuclear reactor breeder blankets for the production of plutonium or uranium-233, high-temperature bearing lubricants, and fuel-reprocessing media. [F.F.B.; R.E.T.]

Methods of investigation. The Gibbs phase rule, a development from established thermodynamic principles, had great consequence in effecting extensive research activities concerning many types of phase equilibrium studies. After its enunciation, F. A. H. Schreinemakers, H. W. B. Roozeboom, N. L. Bowen, and others devised methods for describing heterogeneous phase relationships explicitly and techniques for making quantitative measurements in equilibrium systems. These techniques have served as basic tools to the continuing develop-

General classification. The most familiar type of fuse is the screw-plug fuse used in domestic electric systems (Fig. 1). It consists of a shell, similar to the base of an incandescent lamp; a fusible element connected between the screw shell and the center terminal; and a transparent protective cover.

Other fuses are cartridge fuses in which the fusible element is connected between metal ferrules at either end of an insulating tube (Fig. 2). This tube is usually filled with an insulating material, such as chalk, sand, or a suitable liquid, which cools and quenches the arc that forms when the fusible element melts.

In some types the cartridge is unfilled and open at one end to allow the arc gases to escape to the open atmosphere. These are known as expulsion fuses. Some expulsion fuses and liquid-filled fuses have a spring arranged to pull the terminals of the fusible element apart quickly when it melts, thus rapidly lengthening the arc and expediting its extinction.

There are many ratings and types of cartridge fuses ranging in size from those about 1 in. long, [illegible]

Less commonly used are open link fuses, which are merely strips of fuse material bolted to open terminal blocks.

Operation. The time required for excessive current to melt a fuse varies inversely as the value of the current. However, the nature of this variation, namely the shape of the time-current curve, depends to a marked degree upon the size, shape, and material of the fusible element. The time required to quench the arc after melting depends upon the nature of the arc-quenching material as well as the fusible element. The sum of the melting and arcing times is known as the interrupting time.

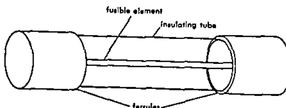


Fig. 2. Cartridge-type fuse.

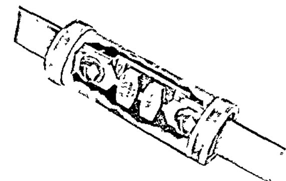


Fig. 3. Renewable-type cartridge fuse, rated 200 amp, 250 volt. (*Bussman Mfg. Division, McGraw Edison Co*)

Special types. When the excess current is high, as when a short circuit occurs on a large power system, the interrupting time of some fuses is so short that the current does not have time to reach its prospective maximum value. Such fuses are known as current-limiting fuses, because they limit the amount of current that actually flows to something substantially less than that which would flow if the interrupting time were longer.

Fuses are frequently used in circuits where it is required that moderately excessive currents, such as during motor starting, be permitted to flow periodically without melting the fuse. Such fuses, called time-delay or time-lag fuses, have relatively longer melting time at moderate overcurrent than those used in circuits where periodic overcurrents do not normally occur.

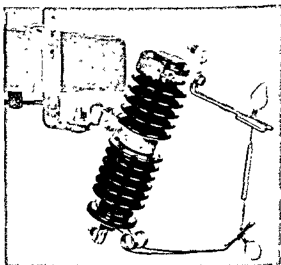


Fig. 4. Flip-open cutout fuse rated 7.8 kv, 50 amp. (*General Electric Co.*)

Some fuses are so constructed that a new fusible element can be inserted to replace one which has been melted. These are known as renewable fuses; an example is shown in Fig. 3.

Many fuses are equipped with a device that gives visual evidence that the fusible element has been melted. These are called indicating fuses.

Fuses used on utility-line poles are sometimes mounted in such a way that when the fuse melts, the fuse holder drops to an open circuit position readily distinguishable from the ground. These are called drop-out fuses. Figure 4 shows a type that springs open when the fuse melts, giving visual indication and providing rapid arc extinction.

A major limitation to the use of fuses arises from the fact that when a polyphase circuit is equipped with fuses a fault will usually cause only one fuse

Fig. 1. The system NaF–ZrF_4 (From C. J. Barton, W. R Grimes, H. Insley, R. E. Moore, and R. E. Thoma, *Phase equilibria in the systems NaF–ZrF_4, UF_4–ZrF_4 and NaF–ZrF_4–UF_4, J. Phys. Chem., 62(6).665–667, 1958*)

Fig. 2 The system NaCl–$ZrCl_4$. (*Oak Ridge Nationar* ...*ies*)

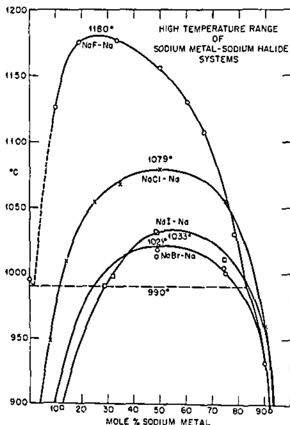


Fig. 3. High-temperature range of sodium metal-sodium halide systems. *(From M. A. Bredig and H. R. Bronstein, J Phys. Chem., 64:64, 1960)*

chloride binary systems with AgCl, CuCl, $FeCl_3$, and $HgCl_2$ all contain immiscible liquids. Ammonium chloride systems with $FeCl_3$ and $HgCl_2$ contain a maximum in the melting curve as well as liquid immiscibility areas. These systems are, how-

tw

so

phases, have been important ever since man learned to recover metals from their compounds either by metallurgical or electrometallurgical processes, particularly molten-salt electrolysis. The solubility of metals in nonmetallic molten phases, especially salts such as the metal halides, can materially affect the yield in such recovery processes. Recent studies of the miscibility of metallic and saltlike liquid phases have not only confirmed a few older observations, but in numerous instances have demonstrated what seemed unexpected in such systems,

liquid metals (cations and electrons) with a common cation. Colloidal suspension was ruled out by the observation of large melting-point depressions and by vapor-pressure measurements. A very high degree of miscibility appears to be confined to metal–metal halide systems of metals having relatively low boiling points, or low cohesive energy, such as the alkali and alkaline-earth metals. However, complete miscibility was also observed in the bismuth–bismuth trichloride system under a chloride vapor pressure of approximately 80 atm.

The molecular and electronic constitution of the solutions appears to be quite different in different systems. The observed rapid increase in electrical conductivity on adding alkali metal to molten alkali halides indicates a different state of the metal in the solution from that of cerium in cerium trichloride, cadmium in cadmium dichloride, and bismuth in bismuth trichloride, where the conductivity of the liquid solution is little different from or actually lower than the (ionic) conductance of the pure molten salt. In the latter cases where mobile electrons are not present, as in the alkali-metal systems, and probably in somewhat similar systems such as nickel in nickel dichloride, gallium in gallium dichloride, and in others, an oxidation-reduction reaction, or the formation of a metal ion of a valency lower than normal, is the likely mechanism for the dissolution of the metal in the salt. In these cases also the solubility of the salt in the metal phase plays a minor part except at high temperatures and salt-vapor pressures. At the high temperature there are also indications that the consitution of the molten-salt phase may undergo a change from a molecular mixture of two salts of the metal, in a normal and in a lower-than-normal oxidation state, to a solution of normal salt and metal ($3Bi_2Cl_2 \rightarrow 2Bi_2 + 2BiCl_3$, M. A. Bredig, 1959).

The interaction between metal and salt leading to miscibility appears to be highly specific, that is, to be confined to a salt and metal with common cation. However, the formation of a subhalide such as GaCl, which, when dissolved in its normal salt, $GaCl_3$, might be mistaken as the solution of the metal in its salt of normal valency, can be enhanced in the presence of a Lewis acid such as $AlCl_3$. This

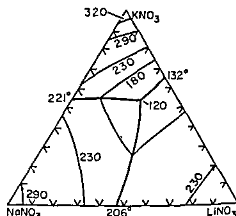


Fig. 4. The system KNO_3–$LiNO_3$–$NaNO_3$. *(From H. R. Carveth, Study of a three-component system, J. Phys. Chem., 2(4):209–228, 1898)*

foreign salt, by depriving the metal ion in its higher, normal valence state, Ga(III), of its complexing and stabilizing halide ions, greatly reduces its stability. [M.A.B.]

Ternary systems. There are few systems of metals or oxides in which such low-melting mixtures occur as in the systems of several salts. Principally for this reason, a marked interest has grown in the industrial application of multicomponent salt systems.

The solid-liquid relationships which occur as interactions of three saltlike compounds having a common anion are similar to those of metals and oxides. In general, the temperatures of the phase changes are lower in the salt systems than in either of the other two types. In each of these three classes of substances, mixtures of three components generally possess lower liquidus values than are to be found in any of the three limiting binary combinations. Addition of further components produces even lower liquidus values, but usually the additional temperature depressions are of small magnitude. Further, the complete determination of the phase equilibria in a single quaternary system requires an understanding of the phase relationships in the six binary and four ternary limiting systems. Few discussions of the results of detailed studies of quaternary systems are to be found.

The number of eutectic points and primary phases in ternary systems is related to the solid-liquid equilibria in the limiting binary systems; thus there is a wide variety in the complexity of phase equilibria in fused-salt ternary systems. Many ternary salt systems are limited by three simple binary systems, each of which contains a eutectic as the single invariant point. A simple three-component system typical of this behavior is that of KNO_3–$LiNO_3$–$NaNO_3$ (Fig. 4). The phase diagram shown here is expressed in the usual form as a polythermal projection of the liquidus surface representing the solid-liquid relationships between the liquidus and the solidus. Solid-state transitions in the limiting binary systems may or may not be manifest in the ternary-system liquidus surface.

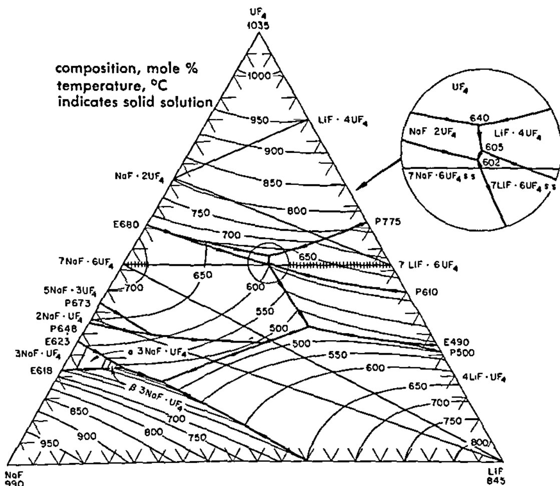


Fig 5. The system NaF–LiF–UF_4. (R. E. Thoma, H. Insley, B. S. Landau, H. A. Friedman and W. R. Grimes, Phase equilibria in the alkali fluoride–uranium tetrafluoride fused salt systems: III, The system NaF–LiF–UF_4, J. Am. Ceramic Soc., 42(1).21–26, 1959)

More complete representation of the equilibria in condensed ternary systems is expressed by isothermal-section diagrams and polythermal diagrams of composition sections.

Several ternary fused-salt systems have been reported in 1958 and 1959, which appear to be as complex as those of the natural minerals systems CaO–MgO–SiO_2 and MgO–SiO_2–Al_2O_3. The collections of phase diagrams published by the American Ceramic Society contain comprehensive surveys of systems reported both here and abroad. Many of the fused-salt phase diagrams are the results of investigations by Russian scientists.

Nearly all the phenomena encountered in the phase behavior of fused-salt systems are exhibited in the phase diagrams of the fluoride systems NaF–LiF–UF_4 (Fig. 5), NaF–ZrF_4–UF_4 (Fig. 6), and NaF–RbF–ZrF_4 (Fig. 7).

The system NaF–LiF–UF_4 contains primary phase fields of eleven solid compounds. Ten invariant points occur within the system. Four of these invariant points involve simply the decomposition or formation of binary compounds. The decomposition on cooling of the compound $3NaF{\cdot}UF_4$ in the ternary system involves pure solids and therefore occurs at the same temperature as in the binary system. The ternary invariant points, at which the binary compounds $5NaF{\cdot}3UF_4$ and $4LiF{\cdot}UF_4$ decompose on cooling, occur at higher temperatures than in the binary systems, because each of these two compounds decomposes on cooling in the ternary system into a pure solid and a solid solution. The binary compound $NaF{\cdot}2UF_4$ is formed in the presence of ternary liquid, not from pure solids as in the binary system, but from UF_4 and $7NaF{\cdot}6UF_4$–$7LiF{\cdot}6UF_4$ solid solution. The temperature of the invariant point is therefore lower than the $NaF{\cdot}2UF_4$ formation temperature in the binary system. The system NaF–LiF–UF_4 contains six compatibility triangles related to two ternary peritectic and four ternary eutectic invariant points.

The principal feature of the system NaF–ZrF_4–UF_4 (Fig. 6) is the existence of continuous solid solutions: that between the compounds UF_4 and ZrF_4 with a minimum at 765°C, that between the congruently melting compounds $7NaF{\cdot}6UF_4$ and $7NaF{\cdot}6ZrF_4$, and that between the congruently melting compounds $3NaF{\cdot}UF_4$ and $3NaF{\cdot}ZrF_4$. The primary phase fields of these solid solutions are the three largest fields in the diagram. Each of the sections $3NaF{\cdot}UF_4$–$3NaF{\cdot}ZrF_4$ and $7NaF{\cdot}6UF_4$–$7NaF{\cdot}6ZrF_4$ constitutes a quasi-binary section of the system. As far as the equilibria involv-

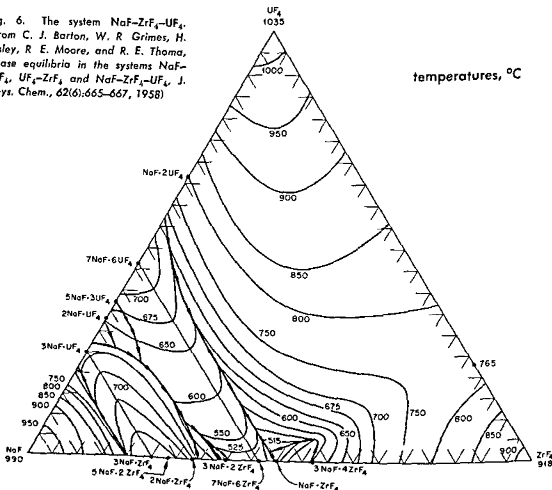


Fig. 6. The system NaF–ZrF_4–UF_4. *(From C. J. Barton, W. R. Grimes, H. Insley, R. E. Moore, and R. E. Thoma, Phase equilibria in the systems NaF–ZrF_4, UF_4–ZrF_4 and NaF–ZrF_4–UF_4, J. Phys. Chem., 62(6):665–667, 1958)*

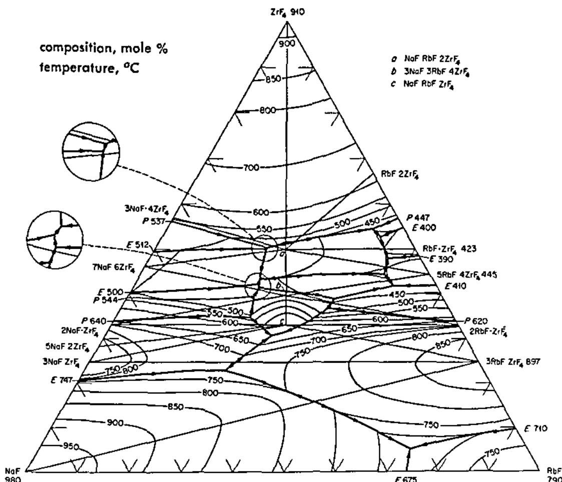


Fig. 7. The system NaF–RbF–ZrF_4. (*Oak Ridge National Laboratories*)

ing the liquidus are concerned, therefore, the system may be divided into three independent subsystems, NaF–3NaF·UF_4–3NaF·ZrF_4, 3NaF·UF_4–7NaF·6UF_4–7NaF·6ZrF_4–3NaF·ZrF_4, and 7NaF·6UF_4–UF_4–ZrF_4–7NaF·6ZrF_4.

The system NaF–RbF–ZrF_4 (Fig. 7) is distinguished among the known fused-salt systems by the large number of primary phase fields, those of some sixteen solid compounds, which it contains. Three of these are of the ternary compounds 3NaF·3RbF·4ZrF_4, NaF·RbF·2ZrF_4, and NaF·RbF·ZrF_4.

Reciprocal systems. Molten-salt systems containing at least two cations and two anions are termed reciprocal systems. In the simplest member of this class of systems, the two cations and two anions can be combined to form four different possible simple salts. The condition of electroneutrality imposes one restriction on the ionic composition of the solution so that the system is a three-component system and can be synthesized by mixing only three of the four possible simple salts.

Reciprocal systems are characterized by the tendency of the two salts which form the stable pair to precipitate in preference to the two salts of the unstable pair. For the alkali halides, E. B. Thomas and L. J. Wood have shown that the stable pair is related to the ionic radii and theoretical lattice energies of the ions in such a way that there is a strong tendency to separate the salt consisting of the small cation and small anion, and the salt containing the large cation and large anion (the two salts which constitute the stable pair). In the phase diagram, this tendency is manifested by liquidus temperatures which are high relative to those of binary systems in which there is a common anion. This is demonstrated in Fig. 8 in which liquidus temperatures for NaF in the NaF–CsCl and NaF–CsBr systems are higher than those in the NaF–ZrF_4 system shown in Fig. 1. In thermodynamic terminology, there are positive deviations from ideal behavior, and in mixtures of the stable components, the escaping tendency of a substance which is a member of the stable pair is higher than if it is a member of the unstable pair.

If the stability of the stable pair relative to that of the unstable pair is high enough, the positive deviations from ideality are manifested not only by relatively high liquidus temperatures but also by the formation of a region of two liquid layers as shown in the CsCl–LiF and CsBr–LiF systems pictured in Fig. 8. In this figure, compositions at temperatures in the region above the horizontal l:

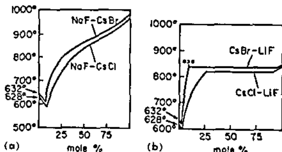


Fig. 8. *(a)* Systems NaF–CsBr and NaF–CsCl. *(b)* Systems CsBr–LiF and CsCl–LiF *(After M. L. Sholokhovich, D. C. Lesnikh, G. A. Bukhalova, and A. G. Bergman, Doklady Akad. Nauk. S.S.S.R., 103:261, 1955)*

will separate into two liquid layers, one being LiF containing a cesium salt and the other being mainly a molten cesium salt containing LiF.

Mixtures of the two salts of the unstable pair are characterized by a tendency toward negative deviations from ideality at compositions consisting mainly of either member of the pair, that is, there will be lower liquidus temperatures for the precipitation of the members of the unstable pair. At compositions with high concentrations of both members of the unstable pair, the members of the stable pair will usually precipitate first. Two of the three possible precipitating phases for any given composition are the two salts of the stable pair.

The formation of complex compounds complicates these phase diagrams. [R.E.T.]

Bibliography: S. T. Bowden, *The Phase Rule and Phase Reactions*, 1938; R. C. Briant and A. M. Weinberg, Molten fluorides as power reactor fuels, *Nuclear Sci. Eng.*, 2(6):797–803, 1957; P. Drossbach, *Elektrochemie Geschmolzener Salze*, 1938; A. Findlay, *The Phase Rule and Its Applications*, 9th ed., 1951; F. J. Ordway, Techniques for growing and mounting small single crystals of refractory compounds, Research Paper 2229, *J. Research Nat. Bur. Standards*, 48(2):152–158, 1952; R. F. Porter, Heterogeneous equilibria and phase diagrams [illegible]

J. R. Ricci, *The Phase Rule and Heterogeneous Equilibrium*, 1951; J. Zernike, *Chemical Phase Theory*, 1955.

Fuselage

A passenger- and cargo-carrying structure to which wings and tail are attached, the complete assembly being an airplane. The fuselage of early airplanes was an uncovered wooden truss, usually braced with steel wire and angles. To decrease the drag of the fuselage, the truss was later covered with doped fabric. Shortly before World War I, welded steel tube came into use for the truss, giving greater rigidity.

Semimonocoque structure. A major advancement in fuselage design came into common use in the 1920s after light-gage sheet aluminum was developed as an aircraft construction material. This material made possible the use of the outer skin to carry all or part of the fuselage loads. The skin carries all of the loads in pure monocoque or shell construction, but this design has not been feasible on many practical aircraft because sheet metal bends easily and does not take compression or column loads well. It has been necessary in most cases to use stiffening members on the inside of the shell, or semimonocoque construction, usually consisting of rings or frames which run circumferentially around the inside of the fuselage and stringers which run lengthwise. One prominent exception to this arrangement is geodetic construction in which the stringers wrap around the fuselage as they run lengthwise. A large number of these stringers is used so that they cross each other at close intervals. This construction is extremely strong, but it has not found much favor with manufacturers because it is costly and difficult to build.

Most semimonocoque or stiffened-shell fuselages use both frames and stringers to support the skin. The stringers take bending loads from the tail surfaces and the aft fuselage. If only a few stringers are used, the skin will contribute little to the fuselage bending strength. However, if there are enough stringers to prevent local buckling of the skin when it is in compression, the skin will add substantially to the bending strength.

The frames receive point loads and distribute them into the skin. Typical point loads are developed at floor attachments in transport planes, at bomb shackles in bombers, and at the engine mounts in single-engine aircraft. Point loads must be applied at the intersection of the plane of the frame and the plane of the skin if their force components are to be resisted along all three axes. If this cannot be done, additional structural members must be incorporated to transfer the loads to such an intersection of planes.

Concentrated loads also require use of a stiffener to transfer them evenly into a thin web. Stringers accomplish this for the outer skin; long angles are usually riveted to the frames for this purpose. Frames also stabilize stringers which are in compression, and thus add to the rigidity of the whole fuselage. Usually frames are located along the sides of cut-outs in the skin such as windows and access doors (Fig. 1). Shear loads carried in the skin concentrate along the edges of these cutouts, especially in the corners. The frames along the side reinforce them and carry most of these concentrated loads. *See* STRESS CONCENTRATION.

Heavy frames and stringers are commonly found near the juncture of the wing and fuselage because this is where the maximum fuselage loads are found. In most cases, the main wing structure passes through, under, or above the main fuselage structure and continues to carry all of the wing loads. The fuselage is attached to the main wing spars through these relatively heavy frames. However, in some designs, the wings have been cut in

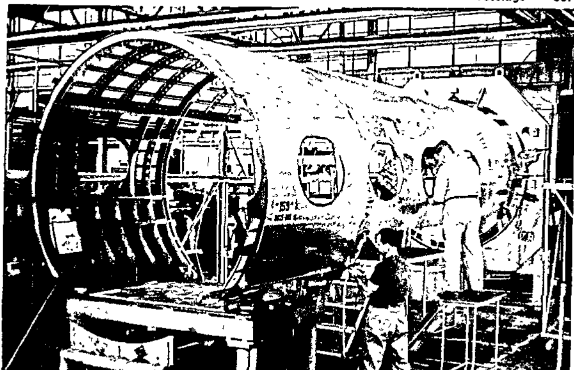


Fig. 1. Fuselage of a turboprop transport under construction is typical of modern semimonocoque fuselage design with frames and stringers spaced at close intervals to reinforce the skin. Windows are oval in shape to avoid stress concentrations which would occur in the corners of square windows. (*Grumman Aircraft Engineering Corp.*)

two with the spars attached to extremely heavy frames which carry the wing loads around the perimeter of the fuselage. This type of wing carry-through structure has been necessary when engines are buried in the wing roots next to the fuselage, and on midwing aircraft which cannot have obstructions such as the wing spars passing through the center of the fuselage. *See* WING STRUCTURE.

Design approach. Many aircraft do not have the most efficient structure with the wing spars going through the fuselage in a straight line; this is a good example of the compromises which must be made in all design. The aeronautical engineer in many instances must subordinate structural simplicity to achievement of external shape that will improve the aircraft's performance, or to a particular arrangement of internal equipment. There is a limit, however, to what can be done structurally, and the aerodynamicist, the power plant engineer, and all of the other designers on an aircraft engineering team must also accept compromises of their original ideas. These involved compromises are endless and are made at every stage of design work.

The job of the structural engineer revolves around one main point: the lightest possible airframe or structure must be built for each aircraft within the restrictions laid down during the design compromises. Two primary methods are available to find the lightest structure to carry any given load. The first is to change the form or geometry of the structure, and for the fuselage designer this usually has meant finding the best arrangement and size for stringers, frames, and skin. The second method is to change the construction material. A fair indication of the complexity of structural design is that most aircraft companies have three or four times more structural engineers than aerodynamicists or any other type of specialist.

Critical loads Detailed structural design begins when information is made available on the exact loads which must be carried. These loads are maneuver, gust, landing, and speed loads The designer must determine which loads will be the most severe for each part of the aircraft. These critical loads are then used to design the part they apply to; the other loading conditions are not investigated thoroughly unless they are nearly equal to the critical load. As an example of this technique, the fuselage of a fighter aircraft is usually designed for a maneuver known as the rolling pullout. This high-stress condition is reached when the aircraft pulls out of a terminal velocity dive and rolls sharply at the same time. Unless the landing gear is housed in the fuselage, an auxiliary rocket engine is carried, or some other unusual feature places concentrated loads into the fuselage, the rolling pullout is usually the design condition for most of the fuselage of fighter planes.

Once the loads have been established, the designer investigates each part in the preliminary structural layout and makes certain that

has the required strength, rigidity, and fatigue life. Such investigations of built-up structures of thin sheet metal are basically similar to all structural investigations, but are usually more involved because of the large number of components among which the loads are distributed. Rigidity in aircraft structures is always of concern because narrow columns and thin sheet have small resistance to compressive loads. *See* BEAM COLUMN.

Fatigue and life. Fatigue failure must be considered because structures which can carry a given load a few times may fail if subjected to that load many times over a long period. Fatigue behavior of metallic materials on the microscopic level is a complicated phenomenon that is incompletely understood. Data regarding the fatigue life of built-up structures are based almost entirely on statistical studies and experiments with typical structural elements.

Effect of aging on the fatigue life of aircraft materials is also largely an unknown quantity. Aging is largely influenced by operational temperatures and vibrational frequencies, as is fatigue resistance. Aging involves changes in the chemical and physical properties of a material, and therefore in its strength and mechanical properties. Statistical studies of parts which have seen long aircraft service is one of the primary methods of studying aging and fatigue. Close inspection of the long-time parts which are still in service is the only means today, however, of ensuring their integrity.

Choice of materials. The structural materials to be used in an aircraft are normally determined early in the design process. Four items generally are of major importance in the selection of materials: strength-to-weight ratio, cost, availability, and ease of fabrication. Sitka spruce, canvas, steel wire, and steel tubing were the main aircraft construction materials until the 1920s. Since that time, aluminum has proved to be an easily worked and fabricated material, and a large aluminum processing industry has grown up, making the metal relatively inexpensive and easy to obtain.

The first successful aluminum alloy for aircraft use was developed in Germany during World War I. Shortly after the war, the alloy, 17S-T, was produced in quantity in the United States and was put into wide use. By the 1930s, continuous research paid off, and a stronger alloy, 24S-T, became commercially available. Most allied aircraft were made of this material during World War II. The Japanese during that war were the first to use an even better alloy which was known in the United States as 75S-T. This alloy had a yield compression stress of about 70,000 psi compared to around 45,000 psi for 24S-T, and it was largely responsible for the light, high-performance Japanese aircraft such as the Zero.

As far as a strength-to-weight comparison is concerned, however, aluminum is better than spruce wood only in tension. Spruce is a much more efficient material in bending and compressive buckling on a weight basis. Aluminum won favor largely because it is more ductile than wood; it will yield slightly under high local stress and will redistribute the load. Under the same conditions, wood and other brittle materials will fail. The ductility of aluminum is almost a necessity for structures with many riveted and bolted connections and many cutouts.

Magnesium has been used extensively in aircraft since the late 1930s. It has a lower density than aluminum, and therefore, follows a pattern similar to that of spruce when compared to aluminum on a strength-to-weight basis. Magnesium is not as efficient for tension loads, but is better in bending and for compression buckling. Principal drawbacks to magnesium have been that it is more susceptible to corrosion and is more difficult to fabricate than aluminum.

The constant search for more efficient structures led to the development of sandwich-type materials with the idea of taking advantage of the best properties of low-density and high-density materials at the same time. The first high-performance aircraft to use sandwich construction extensively was the British Mosquito bomber during World War II. Its sandwich was a thick balsa wood filler or core, bonded to two thin layers of face material which was Sitka spruce. More modern sandwich construction has included aluminum, steel, and titanium face materials with cores of wood, rubber, plastics, steel, and aluminum in the form of honeycombs, trusses, tubes, and webs (Fig. 2). The combinations of face materials and cores which have been experimented with have been almost infinite. Sandwich construction can be compared to steel I beams in that it distributes the major weight of the material to the outer surface where it is the most effective in resisting bending and compression loads.

Sandwich construction can decrease the weight of many major conventional structural elements by 60% or more. The total weight of an aircraft using sandwich materials almost completely has been estimated to be in the neighborhood of 50% less than one using conventional sheet metal construction. Cost, the difficulty of fabrication, and low reliability of the bonding agents between face and core have been the principal objections to the wide use of sandwich materials. These difficulties are being remedied today, and significant improvements have been made in bonding techniques and in the development of large automatic machines for turning out finished sandwich sections.

Pressurized cabin. Probably the most rapid advance in fuselage design came about with the introduction of jet-powered high-altitude passenger transports. But rapid climb to altitude plus internal pressurization caused stress reversals several times, depending on the rate of climb of the airplane and the pumping rate of the pressurization system. Similar stress reversals occurred during descent. As a consequence, small fatigue cracks de-

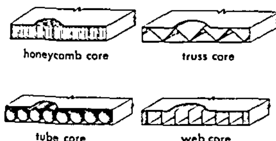


Fig. 2. Sandwich configurations.



Fig. 3. Full-scale jet airliner fuselage is suspended by stub wings inside water tank. Water pressure inside cabin is cycled to simulate changes in altitude and cabin pressurization. Rams above fuselage apply forces to stimulate take-off and landing. (*Convair Division, General Dynamics Corp.*)

veloped around the windows of some early jet transports. These cracks quickly increased to large tears because of high internal pressure needed in the fuselage to keep passengers comfortable at near-sea-level pressure while the airplane cruised at altitudes above 25,000 ft.

Prompt and thorough studies of failures of several jet transports were made available by the British to aircraft manufacturers through the world. The data led to new design methods and standards for the pressurized fuselages of large jet transports. To make sure that the new design methods have been applied properly to each new jetliner, most of them have been subjected to ground tests which simulate the rapid load reversals and flexing that they undergo in flight. This is done by completely submerging the fuselage in a large tank of water (Fig. 3). The necessary stress levels can be produced in the water with only a few pounds per square inch pressure. If air were used for this testing, there would be danger of explosive decompression just as there would be in flight. The wings of the aircraft stick out of the sides of tanks and have the flight loads applied to them with external jacks. Water-tank tests on both sides of the Atlantic have produced important data on the design of fuselages with long fatigue life which have little likelihood of failing catastrophically if they do develop cracks.

Fail-safe design. The new design approach which has been adopted almost universally is generally called the fail-safe system With this system, a designer predicts the fatigue life of his primary structure. He then arranges the structure so that if one or two members fail at once, the entire structure will not collapse. Tears in the skin must be stopped within a short distance by a reinforcing member underneath. If frames or stringers fail, the surrounding structure must be strong enough to carry the total load for a considerable period. One requisite for structures of this type is to provide access to main members so that they may be inspected for cracks before the cracks become too large. Careful inspection, immediate repair of small damage, and fail-safe construction are generally accepted as the most dependable method of maintaining reliable fuselages on large jet aircraft.

Fail-safe design may be considered to be a new method because it introduced a definite new consideration. If this method is used, each part must be examined for strength, stiffness, fatigue, and fail-safe contribution to the structure around it. However, fail-safe design does not alter the general appearance of an aircraft fuselage Some unusual arrangements have been studied experimentally, such as ringing the outside of the fuselage with straps in the same manner that a barrel is held together. This provided an excellent arrangement to keep skin cracks from growing, but it was unacceptable aerodynamically because it left the fuselage with an exceedingly uneven exterior.

In most cases, fail-safe fuselage construction has consisted of moving the frames and stringers closer together and increasing the gage of the skin. The normal safety factor used in aircraft design gave most structures a substantial fail-safe quality, even though they were not specifically investigated for this feature. Only slight modification and strengthening is needed to meet the most stringent fail-safe requirements with an initially sound fuselage design.

High temperature. Aerodynamic heating is the main problem connected with advanced aircraft today. Aluminum construction is not suitable above a speed of about two and one-half times the speed of sound (around 1600 mph). Frictional heat gener-

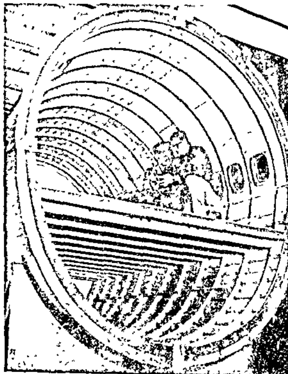




Fig. 4. (a) Specially built 19-ft section of the jet transport fuselage used to check sound levels in the passenger cabin. Ends of the section are sealed; noise is generated electronically in the building housing the section, and measurements are made to determine the effectiveness of the fuselage insulation in damping out the noise. (b) Shown are some of the instruments used to make the measurements and an installation of glass fiber insulation blankets which were tested. (Convair Division, General Dynamics Corp.)

ated as an aircraft moves through the air at 2000 mph (about three times the speed of sound) brings the equilibrium temperature of the outer skin up to about 1000°F. Steel sandwich is to be used as the skin on the Air Force–North American B-70 bomber, a 2000-mph aircraft under development in the United States. Titanium is another high-temperature metal that is in general use today.

A major design problem is that of an aircraft which will fly 18,000 mph and can go into a permanent orbit around the earth above the atmosphere where it will not experience aerodynamic heating. When this aircraft, now under development by the United States and called the Dyna-Soar, returns to earth, the fuselage nose and the leading edges of the wings will have a temperature of at least 3000°F for possibly an hour. Temperatures on the aft sections of the Dyna-Soar will probably remain above 1500°F during this high-temperature period of the reentry.

Two main problems arise in designing very high temperature fuselages. First is the selection of a material that will retain its strength at elevated temperature. Second is the prevention of thermal stresses, which are caused by uneven heating of the structure and can cause it to warp out of shape and finally fail. *See* THERMAL STRESS.

Engineers believe that a Dyna-Soar structure can be built out of materials now in use if insulation and cooling systems are used to cool the heavily stressed members. However, a major research effort is in progress to develop improved materials for high-temperature application, and will undoubtedly continue for many years. Two basic approaches are being followed in this research. One is to disperse stable, heat-resistant particles in an alloy by precipitation hardening. As long as the particles remain intact, they stiffen the metal in much the same manner that steel rods strengthen concrete. As the temperature increases, the particles begin to dissolve into the alloy, and the material loses its stiffness. The other basic approach is to start with base metals for alloys which have very high melting temperatures. Tungsten and columbium, which melt at 6000°F, are being considered for this use, but they have serious drawbacks characteristic of the whole family of high-melting-point metals, the main one being rapid oxidation at high temperatures. Coatings to prevent this wasting away of the base metals are necessary to their eventual use in practical construction.

A number of ingenious approaches to the design of fuselages with expansion joints that essentially eliminate thermal stresses have been proposed by industrial and governmental engineers. Overcoming the mechanical design problem of thermal stresses proved to be much simpler than the basic physical and metallurgical problem of developing high-temperature alloys.

Noise. On jet-powered aircraft, noise is a problem, not only because of passenger discomfort, but also from a structural standpoint. Internal acoustic insulation decreases the noise heard by passengers (Fig. 4). The high-frequency, high-intensity sound waves produced by a jet engine have caused the skin of many aircraft to fatigue and crack after short periods of operation. Four jet engines produce sufficient noise to cause long tears in a conventional fuselage. At full power, they could cause failure in a fuselage designed for use with propellers. Jet aircraft have required some redesign after flight testing because acoustical loads were originally underestimated. Jet-engine noise reduction has been under study for a considerable period. The basic procedure is either to reduce the velocity of the jet exhaust or to spread it out so that there is a larger area of contact between the high-speed air of the jet and the slow-speed air passing over the aircraft (Fig. 5). This velocity differential is responsible for most of the screech from a jet engine. The problem is that if the jet velocity is decreased, the power of the engine is reduced. If the jet exhaust is spread out using a special nozzle, the nozzle usually adds a great deal of drag. Neither the power loss nor the added drag can be tolerated on jet transports which must operate at a profit. Noise reduction apparently will continue to be the subject of research for years to come. *See* AIRCRAFT NOISE; *see also* AIRCRAFT TESTING; AIRFRAME; AIRPLANE. [J.S.BU.]

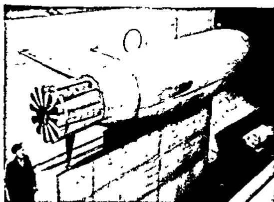


Fig. 5. Typical jet-engine exhaust nozzle to lower engine noise ready for test. Nozzle increases the area of *surface contact between the high-speed jet exhaust* and the surrounding air. (NASA)

Bibliography: M. S. Anderson and R. G. Updegraff, *Some Research Results on Sandwich Structures*, NACA TN 4009, 1957; H. F. Hardrath and F [illegible] s [illegible] [illegible] a [illegible] I [illegible] P. Kuhn and R. W. Peters, *Some Aspects of Fail-Safe Design of Pressurized Fuselages*, NACA TN 4011, 1957; D. J. Peery, *Aircraft Structures*, 1950; J. P. Peterson, *Weight-Strength Studies of Structures Representative of Fuselage Construction*, NACA TN 4114, 1957; R. A. Pride, J. B. Hall, Jr., and M. S. Anderson, *Effects of Rapid Heating on Strength of Airframe Components*, NACA TN 4051, 1957.

Fusion, nuclear

One of the primary nuclear reactions, the name usually designating an energy-releasing rearrangement collision which can occur between various isotopes of low atomic number. *See* NUCLEAR REACTION.

Interest in the nuclear fusion reaction arises from the expectation that it may someday be used to produce useful power, from its role in energy generation in stars, and from its use in the fusion bomb (*see* HYDROGEN BOMB; STELLAR EVOLUTION). Since a primary fusion fuel, deuterium, occurs naturally and is therefore obtainable in virtually inexhaustible supply (by separation of heavy hydrogen from water, 1 atom of deuterium occurring per 5000 atoms of hydrogen), solution of the fusion power problem would permanently solve the problem of the present rapid depletion of chemically valuable fossil fuels. As a power source, the lack of radioactive waste products from the fusion reaction is another argument in its favor as opposed to the fission of uranium. Also, a fusion reactor, by its very nature, could never explode; it could only collapse.

In a nuclear fusion reaction the close collision of two energy-rich nuclei results in a mutual rearrangement of their nucleons (protons and neutrons) to produce two or more reaction products, together with a release of energy. The energy usually appears in the form of kinetic energy of the reaction products, although when energetically allowed, part may be taken up as energy of an excited state of a product nucleus. *In contrast to neutron*-produced nuclear reactions, colliding nuclei, because they are positively charged, require a substantial initial relative kinetic energy to overcome their mutual electrostatic repulsion so that reaction can occur. This required relative energy increases with the nuclear charge Z, so that reactions between low Z nuclei are the easiest to produce. The best known of these are the reactions between the heavy isotopes of hydrogen, deuterium and tritium.

Fusion reactions were discovered in the 1920s when low Z elements were used as targets and bombarded by beams of energetic protons or deuterons. But the nuclear energy released in such bombardments is always microscopic compared with the energy of the impinging beam. This is because most of the energy of the beam particle is dissipated uselessly by ionization and single-particle collisions in the target; only a small fraction of the impinging particles actually produce reactions.

Nuclear fusion reactions can be self-sustaining, however, if they are carried out at a very high temperature. That is to say, if the fusion fuel exists the form of a very hot ionized gas of strip

clei and free electrons, called a "plasma," the agitation energy of the nuclei can overcome their mutual repulsion, causing reactions to occur. This is the mechanism of energy generation in the stars and in the fusion bomb. It is also the method envisaged for the controlled generation of fusion energy, which would be carried out in a very tenuous plasma confined in a so-called magnetic bottle (*see* PLASMA PHYSICS).

The cross sections (effective collisional areas) for many of the simple nuclear fusion reactions have been measured with high precision. It is found that the cross sections generally show broad maxima as a function of energy and have peak values in the general range of 0.01 barn (1 barn = 10^{-24} cm²) to a maximum value of 5 barns, for the deuterium-tritium (D-T) reaction. The energy releases of these reactions can be readily calculated from the mass differences of the initial and final nuclei, or determined by direct measurement.

Simple reactions. The following list gives some of the important simple fusion reactions, their reaction products, and their energy releases in millions of electron volts (Mev).

$$\begin{array}{lll} D + D \rightarrow & He^3 + n + & 3.25 \text{ Mev} \\ D + D \rightarrow & T + p + & 4.0 \text{ Mev} \\ T + D \rightarrow & He^4 + n + & 17.6 \text{ Mev} \\ He^3 + D \rightarrow & He^4 + p + & 18.3 \text{ Mev} \\ Li^6 + D \rightarrow & 2He^4 + & 22.4 \text{ Mev} \\ Li^7 + p \rightarrow & 2He^4 + & 17.3 \text{ Mev} \end{array}$$

If it is remembered that the energy release in the chemical reaction in which hydrogen and oxygen combine to produce a water molecule is about 1 ev per reaction, it will be seen that gram for gram, fusion fuel releases more than one million times as much energy as typical chemical fuels

The two alternative D-D reactions listed occur with about equal probability for the same relative particle energies. Note that the heavy reaction products, tritium and helium-3, may also react, with the release of a large amount of energy. Thus it is possible to visualize a reaction chain in which 6 deuterons are finally converted to 2 helium-4 nuclei, 2 protons, and 2 neutrons, with an over-all energy release of 43 Mev, that is, of about 10^5 kilowatt hours of energy per gram of deuterium. This energy release is several times that released per gram in the fission of uranium, and several million times that released per gram by the combustion of gasoline.

Cross sections. Figure 1 shows the measured values of cross sections as a function of bombarding energy up to 100 kev for the total D-D reaction (both D-Dn and D-Dp), the D-T reaction, and the D-He³ reaction. The most striking feature of these curves is their extremely rapid fall-off with energy as bombarding energies drop to a few kilovolts. This effect arises from the mutual electrostatic repulsion of the nuclei, which prevents them from approaching closely if their relative energy is small (*see* NUCLEAR STRUCTURE). The fact that reactions can occur at all at these energies is attributable to

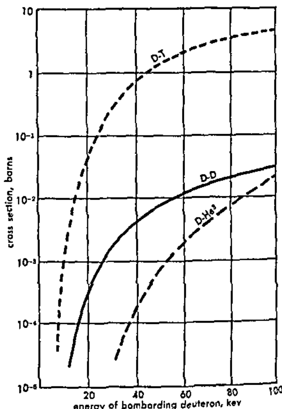


Fig. 1. Cross sections versus bombarding energy for three simple fusion reactions. (*From R. F. Post, Fusion power, Sci. American, 197(6):73–84, 1957*)

the finite range of nuclear interaction forces. In effect, the boundary of the nucleus is not precisely defined by its classical diameter. The role of quantum mechanical effects in nuclear fusion reactions has been treated by G. Gamow and others. It is predicted that the cross sections should obey an exponential law at low energies. This is well borne out in energy regions reasonably far removed from resonances (for example below about 30 kev for the D-T reaction). Over a wide energy range at low energies, the data for the D-D reaction can be accurately fitted by a Gamow curve, the result being (W in kev)

$$\sigma_{D\text{-}D} = \frac{288}{W} e^{-45.8W^{-1/2}} \times 10^{-24} \text{ cm}^2 \qquad (1)$$

for the cross section.

The extreme energy-dependence of this expression can be appreciated by the fact that between 1 and 10 kev, the predicted cross section varies by about thirteen powers of ten, that is, from 3×10^{-43} cm² to 1.5×10^{-29} cm².

Energy division. The kinematics of the fusion reaction stipulates that the reaction can occur only if two or more reaction products result. This is because both mass-energy and momentum balance must be preserved. When there are only two reaction products (which is the case in all of the important reactions) the division of energy between the reaction products is uniquely determined, the

tion's share always going to the lighter particle. The energy division is as follows (disregarding the initial bombarding energy): If

$$A_1 + A_2 \rightarrow A'_1 + A'_2 + Q \qquad (2)$$

with the A's representing the atomic masses of the particles, and Q the total energy released, then

$$W(A'_1) + W(A'_2) = Q \qquad (3)$$

and

$$W(A'_1) = Q\left(\frac{A'_2}{A'_1 + A'_2}\right) \qquad (4)$$

$$W(A'_2) = Q\left(\frac{A'_1}{A'_1 + A'_2}\right) \qquad (5)$$

Thus in the D-T reaction, for example, A'_1, the mass of the α-particle, is four times A'_2, the mass of the neutron, so that the neutron carries off 4/5 of the reaction energy, or 14 Mev.

Reaction rates. When nuclear fusion reactions occur in a high-temperature plasma, the reaction rate per unit volume depends on the particle density of the reacting fuel particles, and on an average of their mutual reaction cross sections and relative velocity v over the particle velocity distributions (*see* THERMONUCLEAR REACTION). For dissimilar reacting nuclei (such as D and T), the reaction rate is

$$R_{12} = n_1 n_2 (\sigma v)_{12} \qquad \text{reactions/(cm}^3\text{) (sec)} \qquad (6)$$

For similar reacting nuclei (for example D and D) it is

$$R_{11} = \tfrac{1}{2} n^2 (\sigma v) \qquad (7)$$

Note that both expressions vary as the *square* of the total particle density (for a given fuel composition).

If the particle velocity distributions are known, (σv) can be determined as a function of energy by

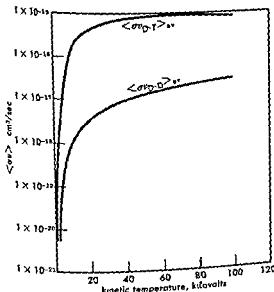


Fig. 2. Plot of $\langle\sigma v\rangle$ versus kinetic temperature for the D-D and D-T reactions.

numerical integration, using the known reaction cross sections. It is customary to assume a maxwellian particle velocity distribution, toward which all others tend in equilibrium. The values of (σv) for the D-D and D-T reactions are shown in Fig. 2. In this plot the kinetic temperature is given in units of kiloelectron volts, 1 kev kinetic temperature = 1.16×10^7 Kelvin degrees. Just as in the case of the cross sections themselves, the most striking feature of these curves is their extremely rapid fall-off with temperature at low temperatures. For example, although at 100 kev for all reactions (σv) is only weakly dependent on temperature, at 1 kev it varies as $T^{6.3}$ and at 0.1 kev as T^{133}! Also, at the lowest temperatures, it can be shown that only the particles in the "tail" of the distribution, which have energies large compared with the average, will make appreciable contributions to the reaction rate, the energy-dependence of σ being so extreme.

Critical temperatures. The nuclear fusion reaction can obviously be self-sustaining only if the rate of loss of energy from the reacting fuel is not greater than the rate of energy generation by fusion reactions. The simplest consequence of this fact is that there will exist critical or ideal ignition temperatures below which a reaction could not sustain itself, even under idealized conditions. In a tenuous plasma, such as would presumably be used in a fusion reactor, ideal or minimum critical temperatures are determined by the unavoidable escape of radiation from the plasma. A minimum value for the radiation emitted from any plasma is that emitted by a pure hydrogenic plasma in the form of x-rays or bremsstrahlung Thus the hydrogen plasma might be expected to possess the lowest ideal ignition temperatures. This is indeed the case: it can be shown by comparison of the nuclear energy release rates with the radiation losses that the critical temperature for the D-T reaction is about 4×10^{7}°K. For the D-D reaction it is about ten times higher. Since both radiation rate and nuclear power vary with the square of the particle density, these critical temperatures are independent of density over the density ranges of interest in plasma physics.

The concept of the critical temperature is a highly idealized one, however, since in any real cases additional losses must be expected to occur which will modify the situation, increasing the required temperature.

Fusion reactor. The nuclear fusion reactor would be a device in which controlled, self-sustaining nuclear fusion reactions would be carried out in order to produce useful power. The reaction would be carried out in a very hot but tenuous fuel gas mixture of hydrogen, helium, or lithium isotopes. To avoid immediate quenching of the reaction, it would have to be carried out in an evacuated chamber, with means provided to prevent the reacting fuel from coming in contact with the chamber walls. The "magnetic bottle" is the only pr foresee ethod of achieving this.

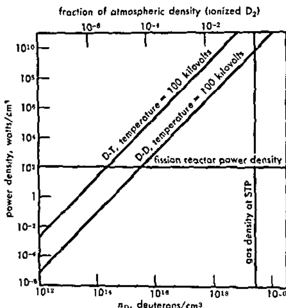


Fig. 3 Power density for D-D and D-T reactions in fusion reactor.

Although nuclear fusion reactors do not yet exist, it is possible to state certain general requirements which such reactors would have to satisfy. These would in turn influence characteristics such as power output and size.

Power density. First it is clear from the preceding discussion that the operating plasma temperature of any fusion reactor would have to lie substantially above the critical temperature defined. This being the case, it is possible to calculate the power density of the reactor, that is, the watts per cubic centimeter released as nuclear power, at some typical operating temperature, as a function of the fuel particle density. A plot of the power density for the D-D and D-T reactions is shown in Fig. 3. In any practical steady-operating power-producing device, engineering limitations on heat transfer will set an upper limit on the power density which can be used. A typical limitation would be 100 watts/cm³ (100 megawatts/m³) as indicated on the plot. The most important consequence of this is that the particle densities required to achieve this power density are very small, that is, 10^{-4}–10^{-5}, compared with ordinary atmospheric gas densities, so that the fuel of a steady-state fusion reactor would be tenuous indeed. It follows that the mean reaction time of the fuel particles will be rather long—1 second or longer—so that the problem of confinement of the reacting fuel for a time long enough to allow reactions to take place is by all odds the most difficult aspect of the controlled fusion problem. Not only does low particle density imply long required confinement time, but also long reaction mean free paths, which might typically be of the order of the earth's circumference.

Note that the plasma heat content U at these densities is rather small, despite the enormous kinetic temperature, because it is proportional to the product of temperature and particle density ($U = \frac{3}{2} nkT$, where k is Boltzmann's constant). At 10^{14} cm^{-3} and 100 kev, $U = \frac{3}{2} \times 10^{14} \times 1.6 \times 10^{-7} = 2.4 \times 10^7$ ergs/cm³, or only about 0.6 cal/cm³. It follows that the energy fluxes required to heat such a plasma should be modest, barring large inefficiencies.

Power balance. The final condition of power balance in a fusion reactor will be determined by a competition between the recoverable electrical power from the nuclear energy released (in the form of energetic reaction products) and *all* of the power required to heat and confine the plasma.

The most important power losses in most visualizations of a reactor are those associated with generating the magnetic field. If one idealizes a fusion [illegible] …nded [illegible] …ious [illegible] …ction is that the nuclear power released per unit length of the reactor should exceed the magnet power. The nuclear power per unit length, p_n, that is released varies as (see Eq. 7)

$$p_n \propto n^2 \langle \sigma v \rangle a^2 \quad \text{watts/cm} \tag{8}$$

where a is the radius of the chamber. On the other hand, the particle density can be related to the strength of the magnetic field B through the parameter

$$\beta = \frac{nkT}{(B^2/8\pi)} \tag{9}$$

so that

$$p_n \propto \beta^2 a^2 B^4 \frac{\langle \sigma v \rangle}{T^2} \tag{10}$$

The magnet power per unit length, p_{mag}, can be shown to be independent of the coil diameter (for a constant ratio of inner and outer diameter), varying as

$$p_{mag} \propto B^2 \bar{\rho} \tag{11}$$

where $\bar{\rho}$ is the effective resistivity of the coil conductor material. Thus for a fixed plasma temperature,

$$\frac{p_n}{p_{mag}} \propto \frac{\beta^2 a^2 B^2 \langle \sigma v \rangle}{\bar{\rho}^2} \tag{12}$$

This ratio must be greater than 1 to achieve power balance. For ordinary copper coils, an evaluation of this expression shows that it can only be satisfied if the product ($\beta a B$) is greater than about 10^6 for the D-T reaction or about 10^7 for the D-D reaction. Since the value of B is likely to be limited by mechanical strength considerations to about 10^5 gauss, and the mean value of β will probably need to be substantially less than 1 to assure hydromagnetic stability, it follows that the reactor radius, a, would have to be relatively large, 100 cm or more, to overcome the coil losses (*see* MAGNETO-

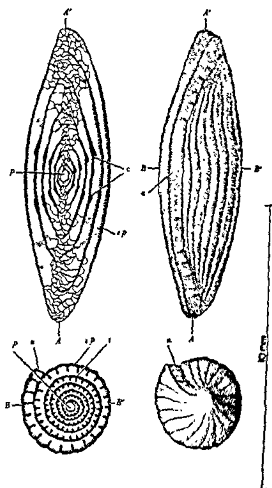
HYDRODYNAMICS). In general, the length of the reactor would have to be great compared with its diameter, so that such a reactor would necessarily be physically large, and necessarily produce a large amount of power. This problem, that is, of large size and overlarge power output, is one of the practical problems which will have to be faced in the future, even after it is shown that fusion reactors are possible in principle. One way in which both reactor size and output might be reduced to more practical values would be to reduce the effective coil resistivity $\bar{\rho}$ by using high-purity coil conductors and refrigerating them to very low temperatures. Over-all reductions in the product ($\beta a B$) of about a factor of five would seem to be possible by use of such cryogenic coils. *See* CRYOGENIC ENGINEERING; PINCH EFFECT. [R.F.P.]

Bibliography: A. S. Bishop, *Project Sherwood*, 1958; S. C. Brown, *Basic Data of Plasma Physics*, 1959; Fusion power—the pace quickens, *Nucleonics*, 15(5):108, 1957; D. B. Hoisington, *Nucleonics Fundamentals*, 1959; R. F. Post, Controlled fusion research, *Revs. Mod. Phys.*, 28(3):338–362, 1956; R. F. Post, Fusion power, *Sci. American*, 197(6): 73–84, 1957; *Proceedings of Second United Nations International Conference on Peaceful Uses of Atomic Energy*, vols. 31–32, 1958; L. Spitzer, Jr., The Stellerator, *Sci. American*, 199(4):28–35, 1958; E. Teller, General problems in the controlled thermonuclear process, *Nuclear Sci. and Eng.*, 1:313–324, 1956.

Fusulinidae

A family of the order Foraminifera which flourished from near the end of the Mississippian to near the end of the Permian. More than 70 genera with more than 1000 distinct species are recognized. Their shells occur in great numbers and are excellent index fossils for relative age determination of rock units. Their classification is based on internal shell features that must be studied from thin sections, one along the axis AA^1 of coiling and through the beginning chamber, and the other at right angles, BB^1. *See* FORAMINIFERA FOSSILS; INDEX FOSSIL.

The fusulinid shell is calcareous and multichambered. All shells started with a spherical chamber; the proloculus *p* and later chambers were added planispirally. As chambers were added, the spirotheca *sp* with small perforations was formed above



A typical fusulinid, *Triticites*, showing positions of thin sections cut along axis of coiling (AA^1) and normal to axis of coiling (BB^1) and showing *a*, antetheca; *c*, chomata; *p*, proloculus; *sp*, spirotheca; and *t*, tunnel

the chambers. The antetheca *a* with minute pores (septal pores) but without an aperture was placed in front of the chamber. All fusulinids tended to reinforce the inside of the chambers with secondary deposits that closed septal pores and wall pores. [illegible] along the margins of the tunnel in the form of small levees, called chomata *c*.

[M.L.T.]